The Impact of Demand Response on Energy Consumption and Economic Welfare *, †, ‡

Manuscript Lasted Updated: November 2, 2023

Maria Jacob
Centre for Net Zero

Ryan Jenkinson
Octopus Energy, KrakenFlex Ltd.

Daniel Lopez Garcia
Centre for Net Zero

Robert D. Metcalfe
Centre for Net Zero, NBER,
University of Southern California

Andrew R. Schein
Centre for Net Zero

Cohen R. Simpson
Centre for Net Zero,
London School of Economics

Lucy Yu
Centre for Net Zero

Abstract: In 2022 and 2023, the Electricity System Operator of Great Britain introduced the Demand Flexibility Service, a program aimed at encouraging utilities and the general population to curtail energy demand during peak times. Octopus Energy, a utility company, implemented its version of the program called Saving Sessions. This initiative comprised 13 individual demand response sessions offered to 1.4 million Octopus Energy customers, with incentives awarded to customers for reducing their energy consumption. Utilizing comprehensive consumer data and employing various identification methods, we estimated the impact of this nationwide program on energy demand and economic welfare. Additionally, we conducted a natural field experiment involving different advance notice periods and incentives provided to customers for their participation in a Saving Session. We found: (i) the Saving Sessions resulted in a 10% reduction in energy demand associated with being invited to participate and, based on an estimated local average treatment effect, a 40% reduction from those actively opting in to Sessions (where individuals manually made adjustments within their homes to change demand); (ii) shorter advance notice periods for signed up customers dampened the demand response from these households by 25%, according to our preferred model specifications; and (iii) the Saving Sessions demonstrated a marginal value of public funds between 1.05 and 2.6, depending on when the grid is approaching a blackout, indicating that the program yielded positive benefits relative to the costs involved.

* **Acknowledgements**: We are very grateful to Lucy Bourke, Sheng Chai, Lukas Franken, Joe Grainger, Andy Hackett, and Izzy Woolgar for their research assistance and formative critique — without which this research would not have been possible. We also thank Kieron Stopforth and Simona Burchill for helping to implement our field experiment, and Rebecca Dibb-Simkin, Greg Jackson, Pete Miller, Alex Schoch, Maddie Inglis, Max McShane, Katrin Weingartner, Sebastian Blake and Marcia Poletti, amongst many others at Octopus Energy Group, for implementing the demand response interventions. We are thankful to Charles Hutchinson and Ben Moll for comments on an earlier draft. AEARCTR-0011100. All mistakes herein are ours alone.

† **Author Contributions**: Authors are arranged in alphabetical order. MJ: analysis, data curation, validation, visualisation; RJ: conceptualisation, analysis; DLG: analysis, project administration, validation, writing (original draft, review, and editing); RDM: conceptualisation, analysis, resources and funding acquisition, validation, project administration, writing (original draft, review, and editing); ARS: conceptualisation, analysis, project administration, writing (original draft, review, and editing); CRS: analysis, project administration, validation, visualisation, writing (original draft, review, and editing); LY: resources and funding acquisition, validation, writing (review and editing).

‡ **Conflicts of Interest**: All authors have a financial stake in the Octopus Energy Group.

1 Introduction

The effective management of energy supply and demand is a major responsibility for grid operators. Indeed, balancing these elementary forces is crucial to preventing blackouts at the level of entire cities, states, and countries. Yet, despite the risk such disruption poses to economic well-being (Allcott et al., 2016, Burgess et al., 2020, Cole et al., 2018, Fisher-Vanden et al., 2015, Gertler et al., 2016), blackouts are frequent occurrences across the world. While their causes vary, contributing factors include uncertainties in energy imports (Fotis et al., 2023, Złotecka and Sroka, 2018), the intermittent nature of renewable energy generation (Carreras et al., 2021, Masood et al., 2018, Yan et al., 2018), and the growing incidence of extreme temperature events (Feng et al., 2022, Jahn et al., 2022, Panteli and Mancarella, 2015).

To clarify how best to manage energy supply and demand, economists have investigated pricing mechanisms and contracts, focusing on, for instance, peak-time pricing (Houthakker, 1951, Joskow, 1976) and interruptible contracts, whereby customers receive payments or pay lower prices for energy in exchange for acceptance of service interruptions during periods of grid constraints (Allcott et al., 2016, Baldick et al., 2006, Tan and Varaiya, 1993). Research on peak-time pricing has been especially plentiful. However, despite numerous smaller-scale evaluations of the impact of peak-time pricing on energy demand (e.g., Caves and Christensen (1980), Caves et al. (1984), Wolak (2007)), there exists no causal evidence around the efficacy of nationwide peak-time pricing campaigns.[1] Here we investigate a nationwide policy to pay consumers to reduce their energy during periods of time when a blackout could occur or carbon-intensive dispatchable energy generation would otherwise be required.

Specifically, we analyzed data from the UK's largest-ever demand flexibility program, which was designed to reduce energy demand at key moments throughout the Winter of 2022-23. This program — known as the Demand Flexibility Service (DFS) — was led by Great Britain's National Grid Electricity System Operator (NGESO) who crafted the program in response to the 2022-23 energy crisis to gauge consumer appetite for reducing electricity demand to help relieve grid pressure and supply scarcity. The DFS involved 22 events taking place from November 2022 to March 2023 during periods of peak energy demand for which British households and businesses were asked to use less electricity.[2] DFS events varied in duration, ranging from one to two hours, and consumer demand reduction was formally remunerated via a price incentive. Specifically, DFS events were organized by NGESO but provided by energy retailers and other aggregators of customers. NGESO gave DFS providers a financial incentive of at least £3,000 for every megawatt hour (MWh) of reduced energy demand.[3] DFS providers were then free to determine how to use their payment from NGESO, allocating some or all of it to incentivizing consumers to reduce demand.

We focused on the 13 DFS events delivered by Octopus Energy — an energy retailer in Great Britain (and other markets) that was the largest of the DFS providers in terms of both the number of participating customers and the level of energy reduction achieved.[4] We investigated whether Octopus Energy's invitations to customers to participate in its variant of NGESO DFS events — branded by Octopus Energy as "Saving Sessions" — impacted customers' electricity demand (in kilowatt hours (kWh)) during events. In so doing, we aimed to better understand the implications of large-scale demand response peak-pricing campaigns to engage in flexible domes-

[1] Research has also explored interventions that are not based on pricing. For example, consider research on conservation appeals and social comparisons during peak consumption periods (Bergquist et al., 2023, Brandon et al., 2019, Ito et al., 2018) in addition to responses to national energy crises such as that of Germany in relation to the Russian-Ukraine war (Moll et al., 2023). There is also a literature on real-time pricing that relates to peak-time pricing (Wolak, 2011).

[2] By peak consumption, we mean energy usage during times when demand on Great Britain's national grid is highest. These periods generally occur from 09:00-11:00 and 16:30-19:30, Monday through Friday.

[3] During DFS events, the marginal unit of energy would have been sourced from a carbon-intensive gas or coal-fired power plant, incurring a marginal private cost of £835/MWh on average for the NGESO, with a maximum cost to NGESO of £5500/MWh.

[4] As of June 2023, Octopus Energy was the third largest domestic electricity supplier in Great Britain, serving 16.9% of GB's domestic electricity market (Ofgem, 2023*c*). Overall, NGESO spent £11M on incentives for the whole Demand Flexibility Service, of which Octopus Energy's Saving Sessions received £6M (National Grid, 2023*d*); for details, see Table AT.30.

tic energy demand.[5] Octopus Energy allocated the majority of its NGESO payment to incentivizing consumers (between £2.25 and £4 per kWh reduced below a customer-specific baseline), allowing us to test how the real-time marginal cost of energy can be used as an incentive for energy demand curtailment.[6]

Causal effects in relation to peak-pricing campaigns are likely to differ between small-scale field experiments and analyses of nationwide data. This difference is because of the totality of the latter. That is to say, during nationwide peak-pricing campaigns, grid operators such as NGESO typically aim to reduce consumption using pricing mechanisms *in addition to an array of state-level communication and political channels*. Thus, *a priori*, we expect that consumer attention and, by extension, consumer behavior will differ in such saturated information environments, and that this difference will yield treatment effects that diverge from those obtained from small-scale experiments (Brewer and Crozier, 2023, Costa and Gerard, 2021, Holladay et al., 2015, Olexsak and Meier, 2014, Prest, 2020, Reiss and White, 2008).[7]

Keeping this distinction in mind, we recovered causal effects in the context of NGESO's peak-pricing campaign by exploiting the structured, two-part nature of consumer engagement with DFS events. Octopus Energy customers were first required to explicitly agree to take part in the overall Saving Sessions campaign (hereafter, *one-time* "sign up"). Once signed-up, Octopus Energy customers were required to explicitly agreed to participate in individual Saving Sessions (hereafter, *event-specific* "opt in" or "Session participation") in response to digital appeals (hereafter, "opt-in notices") sent to customers by Octopus Energy. Opt-in notices typically: (a) communicated the price incentive associated with participation in a specific session; and (b) provided customers with a hyperlink through which they could opt in to a specific session. In general, opt-in notices took the form of emails to Octopus Energy account holders and/or notifications to the account holder via the Octopus Energy mobile application.[8]

We used three sources of variation to construct comparisons between observed and counterfactual outcomes for our response variables of interest — i.e., in-Session electricity consumption (kWh). We formally tested these comparisons using three difference-in-differences (DiD) designs and data from all 13 Saving Sessions. First, the manner in which Saving Sessions were delivered allowed us to contrast the behavior of ≈332k Octopus Energy customers who signed up before the first Saving Session on November 15, 2022 to ≈654k Octopus Energy customers who *never* signed up to take part in DFS events. Second, we compared the behavior of ≈332k Octopus Energy customers who signed up before the first Saving Session to ≈12k Octopus customers who signed up after the ninth Saving Session on January 30, 2023.[9] Third, owing to Octopus Energy's December 2022 acquisition of the British energy retailer Bulb Energy, we compared the behavior of ≈1.137m Octopus Energy customers *invited to sign up to DFS events* to ≈198k newly-acquired Bulb customers who would have been eligible for a sign-up invitation *but who were not invited* due to Bulb not being a DFS provider and due to Bulb's acquisition by Octopus taking place too late in 2022 for Octopus to invite Bulb customers. Thus, Bulb customers provided us with a natural counterfactual

[5] All of the treatments had a financial component, as customers were encouraged to take part in Saving Sessions using economic rewards (see Table 1 and Sections 2.2 and 2.3) — and a portion of our analysis focuses on the causal impact of a "bonus" cash incentive on top of the typical reward rate (Section 4.4). In this respect, all of our treatments hybridized a text-based appeal (Bergquist et al., 2023) with a financial incentive, where the digital appeal serves as the vehicle through which the financial incentive is offered (e.g., see Figures 1 and 2).

[6] For context, the marginal price per kWh for the average Octopus smart meter customer was £0.34/kWh. The 2022-23 Saving Sessions occurred during a time when electricity and gas prices were regulated by the UK Government's Energy Price Guarantee (UK GOV, 2023*b*).

[7] Conducting a field experiment under such information saturation requires careful selection of counterfactual cases in view of credibly identifying causal effects. This is because the pervasiveness of energy-conservation messaging could contaminate the control group or result in violation of the stable unit treatment value assumption (SUTVA). Expected divergence between small-scale experiments and nationwide campaigns also reflects findings from research on scaling and the external validity of experiments in fields beyond energy (Al-Ubaydli et al., 2023, List, 2022).

[8] Customers could receive multiple opt-in notices across the same or different channel (e.g., email, "push" notification to one's mobile device such as a phone or tablet). However, the vast majority of Octopus Energy customers in the data we analyzed received one opt-in notice via email or one email-based opt-in notice alongside one push-based opt-in notice.

[9] We did not see meaningful differences in treatment effects under different definitions of late joining (see Section AT.4 for a comparison of results using our primary model specification and an alternative).

group.[10] In summary, we used DiDs to make three comparisons: (a) Signed Up Early versus Never Signed Up; (b) Signed Up Early versus Signed Up Late; and (c) Octopus Energy customers invited to sign up versus newly-acquired smart-meter Bulb Energy customers not invited to do so.

In judging the validity of DiD in this situation, it is important to note that prior to receiving the invitation to sign up to participate in Saving Sessions, Octopus Energy customers were unaware of the timing of Saving Sessions — in terms of both on which days and at what times Saving Sessions would occur. This minimized the possibility of selection bias around sign-up (e.g., customers who planned to use less electricity during events being more likely to sign up). Indeed, in comparing trends in customers' average consumption during the periods before the start of the Saving Sessions campaign during the month of October and the first half of November, we observed parallel trends for each pair of groups used for our three DiD designs.

Furthermore, we probed important parameters of the design of the provision of Saving Sessions with an eye to how such design might be optimized by governments, grid operators such as NGESO, and retailers like Octopus Energy. To do so, we focused on the behavior of Octopus Energy customers during two unique Saving Sessions — on February 13, 2023 from 17:30 to 18:30, and on March 15, 2023 from 16:30 to 17:30. We do so as the customer-messaging set-ups for these two Saving Sessions involved incentives and/or notice periods that differed in a manner that allowed us to credibly establish causal effects in relation to the broad timing (i.e., When?) of Octopus Energy's appeals to its customers to flexibly use electricity and, for one of our secondary analyses, the general channel through which these appeals were made (i.e., their "type"; e.g., email versus SMS). This timing parameter is especially important to understand as grid operators can choose when to call for a demand response event. Our analysis is, to the best of our knowledge, the first to analyze this timing parameter at a scale expected to have enough statistical power to estimate hour-specific effects.

With regard to experimental design, opt-in notices were typically sent to Octopus Energy customers on the day prior to a given Saving Session (hereafter, "day-ahead" notices). However, we used a regression discontinuity design (RDD) to exploit a technical fault that resulted in both the delay and the time-ordered delivery of opt-in notices for the February 13, 2023 Saving Session in a fashion broadly reflective of customers' "tenure" with Octopus Energy (i.e., the length of time a customer has used Octopus Energy as service provider). Practically speaking, this allowed us to gauge the effect of receiving notice *on the day of the Saving Session itself* or, rather, under "short" notice (hereafter, "day-of" or "intraday" notices) using data from ≈80-350k customers.

For the Saving Session on March 15, 2023, we instead leveraged a field experiment. For this field experiment, we created exogenous variation in the broad time and general type of *supplementary* opt-in notice that were given alongside a universal, primary day-of notice to ≈650k DFS-participating Octopus Energy customers on the morning of March 15, 2023. In particular, we sent a randomly selected group of customers an ancillary "heads-up" email stating that there "may be a Saving Session tomorrow evening". Furthermore, we sent a second, randomly selected group of customers a day-of "reminder" SMS text message while simultaneously making this second group eligible for a bonus price incentive of £1.25.

Across our three experimental designs, we obtained results of substantial scientific and policy relevance. In particular, we showed, using our DiD design involving newly acquired Bulb customers, that *merely being invited to sign up to DFS events* caused a ≈10% reduction in consumption during the half-hours of the 13 Saving Sessions. This effect is a kind of intent-to-treat effect (ITT) interpretable as the average change in consumption amongst smart-meter customers during DFS events when their *energy supplier* (here, Octopus Energy) participates in the DFS. With regard to the impact of taking part in DFS events, our DiD designs indicated that *signing-up* to Saving Sessions caused ≈24-27% lower consumption during Saving Session half-hours. Further still, our DiD designs in-

[10] Merger and acquisition designs using consumer data are an underutilized means of obtaining counterfactual cases as accessing consumer data has historically been difficult. For excellent examples of this analytic strategy, see Farronato et al. (2023), Li and Agarwal (2017), Yan et al. (2021).

dicated that Session *opt-in* caused ≈38-44% lower consumption during Saving Session half-hours.[11] We compared our DiD-based results to estimates of reduced consumption obtained using NGESO's preferred methodology for assessing demand reduction — a kind of within-person pre-post comparison. We found that the official NGESO methodology is too optimistic in that it *overestimates* demand reduction by ≈13%.

Beyond consumption, we obtained several noteworthy results related to the probability of Saving Session sign-up and the probability of Session-specific participation — where these results pointed to which customers might be targeted and how peak-pricing campaigns might be designed to maximize engagement.[12] For instance, customers who signed up to Saving Sessions were more likely to have *smart tariffs* (i.e., special energy products for customers with low-carbon technologies like electric vehicles, batteries, and heat pumps) and to live in geographic areas with lower levels of socio-economic deprivation. However, we found no evidence to suggest that probability of sign-up is associated with home energy-efficiency rating, annual energy consumption, or UK region of residence. Conditional on sign-up, we obtain similar results for Session opt-in, where Octopus Energy customers are more likely to opt-in when they have a smart tariff and live in lower-deprivation areas.

Moreover, we found several corollary results with respect to treatment effect heterogeneity in relation to our DiD designs. In general, we found that all customer sub-groups conserved energy during the Saving Sessions, but customers from lower-deprivation areas, with higher annual consumption, and on smart tariffs all had greater demand reduction. However, even the groups with lower treatment effects showed large, economically meaningful effects of participating in Saving Sessions. For example, the Conditional Average Treatment Effect (CATE) of signing up to Saving Sessions for customers in very low deprivation postcodes was -0.1059 kWh (-22.7% of during-Session consumption of non-signed-up customers from very low deprivation postcodes), while the CATE for customers in very high deprivation postcodes was -0.0644 kWh (-17.2% of during-Session consumption of non-signed-up customers from very high deprivation postcodes). This difference was meaningful, but even the CATE for very high deprivation postcodes was substantial. We also found suggestive evidence that the treatment effect declines over the course of the DFS season. However, with only 13 Saving Sessions, this evidence should be treated with caution.

Our RDD and our field experiment provided evidence to suggest that longer notice period causes greater demand reduction during Saving Sessions. Specifically, our RDD indicated that being sent an intraday notice as opposed to day-ahead notice (i.e., having shorter notice) increased in-Session consumption by 10.6% or, when adjusting for pre-treatment covariates, 7.1%. Likewise, our field experiment indicated that receiving an earlier, ancillary "heads-up" email (i.e., having advanced notice) reduced in-Session consumption by 3.2% or, when adjusting for pre-treatment covariates, 1.6%. The result from our RDD is equivalent to ≈25% of the demand reduction our DiDs identify for the 10th Saving Session, depending on the model specification. The result from our field experiment was equivalent to ≈7% of the demand reduction our DiDs identify for the 12th Saving Session. Thus, we found that a shorter notice period dampened customers' demand response, though that response was still substantial.

Further still, we used a variant of our RDD to estimate treatment effects that were specific to each of the five hours (08:00 to 13:00) during which intraday notices for the 10th Saving Session were sent. We found that every

[11] In terms of behavioral mechanisms, we supplemented our causal analyses using data from surveys sent to 60,000 Octopus Energy customers who had signed up to participate in DFS events (Section AI.2). We found that 70% of respondents adjusted their domestic energy behavior by turning off lights, 57% adjusted their behavior by avoiding use of television, games consoles, and/or other entertainment appliances, 50% shifted the time during which they use their dryer, washer-dryer, and/or washing machine, and 50% eschewed the charging of small appliances. Only 9% of respondents said that they had already planned to be away from their home during Saving Sessions; these customers would be considered inframarginal in our estimates (i.e., receiving compensation for demand reduction that the Saving Sessions did not actually cause). Surveyed customers report high levels of satisfaction with Saving Sessions — including under tweaked designs involving shorter notice periods or payments for using more electricity when the grid had excess supply. Furthermore, surveyed customers report a high willingness to participate in future Saving Sessions. They also disclose that they engaged in an array of behaviors to manually deliver flexibility.

[12] We viewed both forms of participation as indicators of customers' willingness to engage in electricity-conserving behaviors within their home.

hour closer to the start of the 10th Saving Session (17:30) that notice was given, in-Session demand reduction was lower by ≈10%, suggesting a notice period elasticity of demand of 0.7 to 0.8 (i.e., every 1% increase in wait time of the peak-period notice to customers, the conservation effect size in the peak period decreases by 0.7 to 0.8%). Our paper is, to the best of our knowledge, the first to estimate such an elasticity for any energy market.

Evidence around Session participation and notice timing from our RDD was somewhat mixed, although we did find that longer notice period increases the probability of opting into the 10th Saving Session, conditional on pre-treatment variables. Furthermore, our field experiment indicated that being sent a supplementary "heads-up" email the day prior or a day-of "reminder" SMS text message respectively increased the probability of Session participation by ≈6% and ≈23% over baseline — where we observe an increase of ≈5.2% in the probability of Session participation when considering mere eligibility for the SMS-based treatment.

Given that we estimated changes in energy demand and the cost of a policy designed to shape demand response (i.e, the DFS), we also derived associated impacts on societal well-being. In particular, our economic welfare analysis indicated that the DFS was welfare-enhancing when considering the payments from NGESO as a pure benefit to Octopus Energy and its customers. If only considering the reduction in greenhouse gas emissions caused by the displacement of the most expensive (and thus likely marginal) generation on the grid during the Saving Sessions half hours, the marginal value of public funds (MVPF) (Hendren and Sprung-Keyser, 2020) of the Saving Sessions component of the broader DFS program was 1.05. The MVPF was substantially higher (2.63) if one assumes that the demand reduction we identify should be valued according to these environmental benefits *and* at the value that the UK Government ascribes to lost load. In summary, the welfare impacts were sensitive to the extent to which the demand response reduced the likelihood of lost load; note that an MVPF of 2.6 is larger than many other popular policy programs, such as housing vouchers, job training, cash transfers, and adult health subsidies (Hendren and Sprung-Keyser, 2020).

From the perspective of energy grid management, our analyses point to a tension between the size and value of flexibility, at least for domestic consumers. Grid operators such as NGESO likely find it more difficult to forecast lost load for periods further in the future. If it is correct that lost load becomes more certain when the hour in question approaches, results from our RDD and field experiment suggest a potential trade-off. If the notice period given to customers is shorter, the flexibility is more valuable to NGESO and similar grid operators. However, our RDD and field experiment results showed that domestic customers' demand reduction was smaller, though still substantial, when the notice they received was closer to the time of flexibility "delivery".

The paper proceeds as follows. First, we summarize the creation of the DFS by NGESO in Great Britain, including the methodology NGESO prescribes for DFS providers to use to calculate demand reduction and our alternative methodology to measured demand reduction — i.e., difference-in-differences (Section 2). We follow immediately with results obtained when using these two methods to estimate demand reduction (Section 3). We then provide a similarly-structured presentation of methods and results for our RDD and our field experiment (Section 4). Next, we explore the macro-level welfare implications of the 13 Saving Sessions (Section 5). We conclude by outlining the implications of our findings for policymakers and for future research in relation to the limitations of our analysis (Section 6).

2 Background

2.1 Great Britain's Electricity Market

Great Britain's electricity system consists of three interconnected components: (a) electricity generation; (b) high-voltage transmission and low-voltage distribution; and (c) consumption by end-users (National Grid, 2023*f*). At present, the regulatory framework for Great Britain's electricity system involves wholesale markets operating at the national level, encompassing the entirety of England, Scotland, and Wales, rather than exhibiting regional or nodal variations.[13] Zonal considerations, such as constraints on electricity transmission between different regions, are managed by NGESO and are thus beyond the reach of wholesale market dynamics. Indeed, many of the balancing services offered by NGESO involve the management of constrained electricity transmission between different geographic areas. Nevertheless, in the future, Great Britain's energy system could see zonal or nodal pricing (Ofgem, 2022*b*).

Historically, electricity generation in Great Britain mostly relied on fossil fuels, particularly coal and gas. Power plants utilizing these sources of energy possess a "flexible" characteristic, enabling rapid adjustments in output to accommodate fluctuations in demand. As renewable generation capacity has expanded, there has been a growing tendency to draw on the flexibility of gas power plants to balance supply and demand during periods when renewable generation falls short.

Electricity wholesale prices in Great Britain are determined by energy generators' *merit order* (Institute for government, 2022). This approach arranges generation units in ascending order based on their cost — ranging from the least expensive to the most expensive. In the merit order, the most expensive power plant required to meet demand sets the price for electricity. Interestingly, because power generators have some awareness of the market status, they tend to submit bids close to the marginal cost associated with the most expensive plant that ultimately defines the price in the merit order (Grubb et al., 2022, Zakeri et al., 2022), reminiscent of strategic bidding behavior by the dominant electricity generators in England and Wales in the early 1990s (Wolak and Patrick, 2001). Due to the higher operational costs of coal and gas plants, generators of this type frequently emerge as the marginal fuel in Great Britain's merit order. Thus, despite gas generators contributing only around 40% of Great Britain's total energy, gas prices have determined electricity prices about 84% of the time (Grubb et al., 2022).

This dynamic underscores the significance of gas power plants in shaping electricity prices within the wholesale market. Nevertheless, as renewable energy sources continue to make up a larger portion of the energy mix in Great Britain, as in many other countries, the task of balancing energy supply and demand will become increasingly complex. This presents intriguing economic considerations, where the present research focuses on optimal policies for facilitating flexible energy demand in relation to a peak-pricing campaign.

2.2 The Demand Flexibility Service (DFS)

NGESO launched the DFS in November of 2022 to gauge the feasibility of strategically shifting household and business consumption during times of high demand and low supply. The ultimate aim of the DFS was to mitigate the risk of lost load (e.g., brownouts or blackouts) and the need to use carbon-intensive methods of power generation (e.g., gas and backup coal) to avoid electricity shortfalls. Participation in the DFS was open to all households and businesses in Great Britain that had a smart meter providing reliable readings and had a relationship with a DFS provider. DFS providers included not only traditional energy suppliers like Octopus Energy but also asset aggregators and other firms with the capability to monitor and manage consumption (e.g., Loop, a smart-meter-connected energy-saving advice application).

The DFS was, in some respects, a blunt policy tool to encourage national demand response in the face of grid constraints. In the future, we expect consumer flexibility to be more dynamic, automated and routine, through low-carbon technologies optimizing demand in an increasingly digitized grid. British energy retailers do not yet settle their procurement at the half-hourly level for individual customers - most consumers continue to be settled on a 'non-half-hourly' basis using daily reads or usage estimates. And half-hourly wholesale prices are determined at national-level, as discussed in Section 2.1. For these reasons, retailers do not face a direct incentive to encourage

[13] Northern Ireland is independent in terms of its electricity system. The island of Ireland, comprising both the Republic of Ireland and Northern Ireland, maintains an integrated electricity system under the direction of the transmission systems operator EirGrid that is separate from NGESO. Ireland and Great Britain's grids are joined by way of Wales via the Moyle and East–West Interconnectors.

their own customers to reduce electricity demand in half-hours and locations when and where it is scarce. Ofgem aims to implement market-wide half hourly settlement by the end of 2026 (Ofgem, 2023*a*), while the government's Review of Electricity Market Arrangements is considering options to introduce sharper locational signals into the wholesale market. Such reforms would help electricity market price signals to reflect grid optimization needs, and these price signals would likely encourage efficient and innovative arrangements by retailers to encourage demand-response among their customers (Wolak, 2019). British energy retailers might even develop programs like "Saving Sessions" on their own, alongside dynamic tariffs and other demand response tools, to reduce electricity procurement costs in the most constrained half-hours and locations. As Great Britain transitions to this more flexible and dynamic system, the DFS might be viewed as a limited policy response in the interim to help scale demand response and as a grid management contingency.

Consumer participation in the DFS occurred in two stages: (a) one-time *sign-up* to take part in the DFS itself; and (b) repeated, event-specific *opt-in*. NGESO mandated this two-stage process. After they signed up or opted in, participants were not penalized for failing to reduce demand, or even having negative demand reduction (i.e., actual consumption higher than baseline). Even though the DFS was free for consumers, opt-in rates were not 100% — participation rates range from 42-69% among consumers who had signed up for Saving Sessions (Table 1). We further discuss sign-up and opt-in in Section AI.1.

Table 1: Summary of times, incentives, sign-ups and opt-ins for each Saving Session.

Session date	Session Type	Session Start	Session End	Incentive (£/kWh)	Signed Up by Session Start	Opted In to Session	% of Signed-up that Opted In
November 15, 2022	Test	17:00	18:00	2.25	408,925	281,952	68.9
November 22, 2022	Test	17:30	18:30	2.25	427,114	267,514	62.6
November 30, 2022	Test	17:30	18:30	2.25	445,695	273,988	61.5
December 1, 2022	Test	17:00	18:00	2.25	449,167	269,339	60.0
December 12, 2022	Test	17:00	19:00	2.25	468,379	306,869	65.5
January 19, 2023	Test	9:00	10:00	2.25	517,910	286,337	55.3
January 23, 2023	Live	17:00	18:00	3.38	605,223	358,323	59.2
January 24, 2023	Live	16:30	18:00	4.00	615,061	343,458	55.8
January 30, 2023	Test	9:00	10:00	2.25	627,269	314,494	50.1
February 13, 2023	Test	17:30	18:30	2.25	640,892	354,682	55.3
February 21, 2023	Test	17:30	18:30	2.25	658,283	395,946	60.1
March 15, 2023	Test	18:30	19:30	2.25	684,534	289,490	42.3
March 23, 2023	Test	18:30	19:30	2.25	692,534	382,857	55.3

Note: Dates of the 13 DFS events delivered by Octopus Energy throughout Winter 2022-23. Price incentives for Octopus Energy customers were administered via a points-based rewards scheme (i.e., "OctoPoints"), which could be exchanged for cash or account credit. Summary of information for each Saving Session includes each event's: date, time, calculated monetary incentive per kWh of demand reduction, number of customers who signed up for Saving Sessions by the date of the Saving Session, number of customers who opted in to the Session, and the percentage of signed up customers who had opted in.

NGESO organized a total of 22 DFS events between November 2022 and March 2023, each with a duration of one, one and a half, or two hours. Events were divided into two categories — i.e., "test" events and "live" events. DFS providers were required to first deliver two test events (National Grid, 2023*b*).[14] There were only two live events across the Winter 2022-23 period, and they occurred on January 23 and 24, 2023.

During test events, NGESO established a Guaranteed Acceptance Price of £3,000/MWh for all bids submitted by DFS providers to provide demand reduction. This minimum price was intended to offer assurance to DFS providers in the market (National Grid, 2022*b*). Providers had the option to submit bids below the Guaranteed Acceptance Price, which could be accepted at the reduced price. Alternatively, they could present bids above this

[14] In general, only two test events took place each month. However, Octopus Energy delivered three DFS events in November 2022. This is because each DFS provider's initial two compulsory events did not count against their two monthly test events (National Grid, 2023*i*).

price, but they would face the risk of their bids not being accepted if the marginal price in the Balancing Mechanism (the main market used by NGESO for balancing services) turned out to be lower than their bid (Ofgem, 2022*a*).

For live DFS events, the dynamics were distinct. NGESO designed live events to be initiated only after all other grid balancing services and market mechanisms had been exhausted for the relevant time periods. During live events, DFS providers participated in an auction-style mechanism where they submitted bids to reduce as much electricity demand as possible from their consumer base. Unlike the test events where the Guaranteed Acceptance Price set a baseline, the live events allowed DFS providers to present bids at prices higher than the Guaranteed Acceptance Price. This approach was taken to incentivize DFS providers to offer more substantial demand reductions during times when grid balancing was challenging, reflecting the unique circumstances of live events.

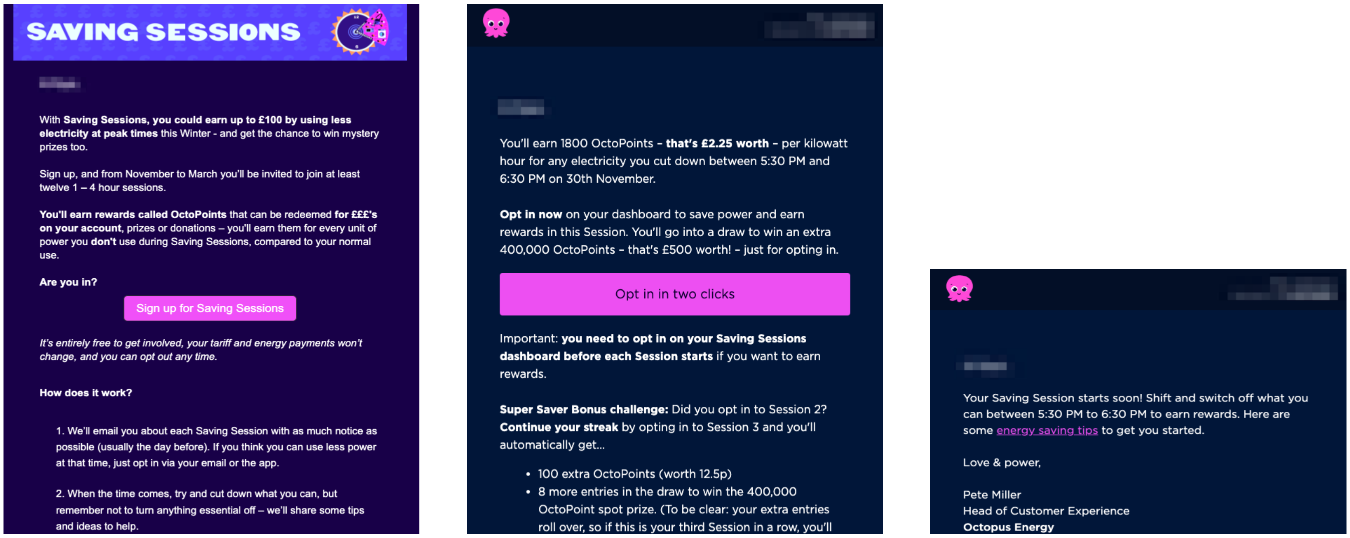
For the two live events on January 23 and 24, 2023, DFS providers submitted a range of bids at prices exceeding the Guaranteed Acceptance Price. While the precise accepted bid prices were not disclosed publicly, information from reporting by NGESO indicates that submitted offers fell within a range of £4,400/MWh and £6,500/MWh (LCP Delta, 2023).

2.3 Design of Treatment: Saving Sessions

Each DFS provider delivered their own variant of the Demand Flexibility Service, often in a special campaign. Octopus Energy branded its DFS implementation “Saving Sessions”.

To market Saving Sessions, Octopus Energy created a dedicated web page where its customers could agree to participate in the DFS (i.e., *one-time* “sign-up”). Octopus Energy invited 1,384,400 customers to sign up (see Figure AF.11). Some customers who were eligible to sign up, such as those with open complaints, were not emailed. Among the invited customers, 375,026 signed up before the first Session, 163,502 signed up between the first Session and the end of January, and 15,643 signed up between the start of February and the final Saving Session; 830,229 did not sign up. In addition, there were 138,363 customers who signed up who were *not* explicitly invited to sign up – for example, because they had an open complaint at the time of the invitations being sent out, or joined Octopus Energy after the invitations had been sent. Thus a majority of invited customers signed up before the first Session on November 15, 2022, after Octopus Energy sent the first sign-up invitation email (Figure 1a).

Figure 1: Saving Session sign up email sent before November 2022 and example email-based opt-in notice sent for the November 30 Saving Session alongside the remind email sent on the day of the event.



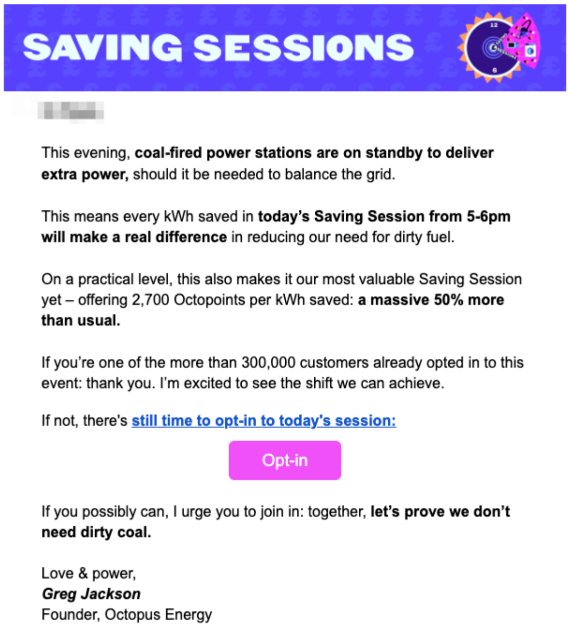
(**a**) Saving Sessions sign up email. (**b**) Initial Opt-in Notice. (**c**) Saving Session Reminder.

Note: In late October 2022, Octopus Energy sent an invitation email to 1.3m customers inviting them to participate in Saving Sessions. This email (left): (a) stated one's potential cash savings (i.e., "up to £100"); (b) noted that there would be a minimum of 12 Sessions over the campaign; and (c) discussed "OctoPoints", a synthetic currency used by Octopus Energy to reward customers for reducing their demand during Sessions. Opt-in email notices were typically sent to customers the day before (center) and reminder emails during the day of (right) a Saving Session. These sample emails were sent for the sessions that took place on November 30 2022, which occurred between 5:30 pm to 6:30 pm. Each kWh of demand reduction was rewarded with 1800 OctoPoints, equivalent to £2.25 per kWh.

To secure agreement to participate in each Saving Session (i.e., event-specific "opt-in"; Figure 1b), Octopus Energy sent emails to customers who had signed up for its DFS events. Signed-up customers with the Octopus Energy mobile application also received opt-in notice in the form of push notification. Opt-in notices (i.e., email and push) were normally sent the day before a Session (i.e., day-ahead notice). However, recall that we also consider two interesting deviations from to this typical messaging set-up wherein customers: (a) received day-of opt-in notice (Section 4.1); or (b) received supplementary opt-in notice (Section 4.4). Regardless of set-up, on the day of each Saving Session, a few hours before the start of the event, Octopus Energy sent customers *who had* opted in an email-based reminder (Figure 1c). For each Saving Session, customers who had opted in were remunerated with "OctoPoints" after each Session.[15]

[15]OctoPoints could be used to reduce one's energy bill, donated to charities, or exchanged for cash. Customers were rewarded with OctoPoints in proportion to their *clipped* (i.e., zero-truncated) demand reduction. Briefly, each OctoPoint was worth £0.00125 — this is 1/800 of £1 — where the value of 800 was chosen as the divisor as octopuses have eight legs. For each kWh of clipped demand reduction in *test* sessions, customers received 1800 OctoPoints — where 1800 OctoPoints/kWh is equivalent to £2.25/kWh. In Section 2.4.1, we further discuss clipping and how DFS providers calculated demand reduction.

Figure 2: Extra Email-based Notices for the Saving Sessions on 23 and 24 January, 2023.



SAVING SESSIONS

This evening, **coal-fired power stations are on standby to deliver extra power,** should it be needed to balance the grid.

This means every kWh saved in **today's Saving Session from 5-6pm will make a real difference** in reducing our need for dirty fuel.

On a practical level, this also makes it our most valuable Saving Session yet – offering 2,700 Octopoints per kWh saved: **a massive 50% more than usual.**

If you're one of the more than 300,000 customers already opted in to this event: thank you. I'm excited to see the shift we can achieve.

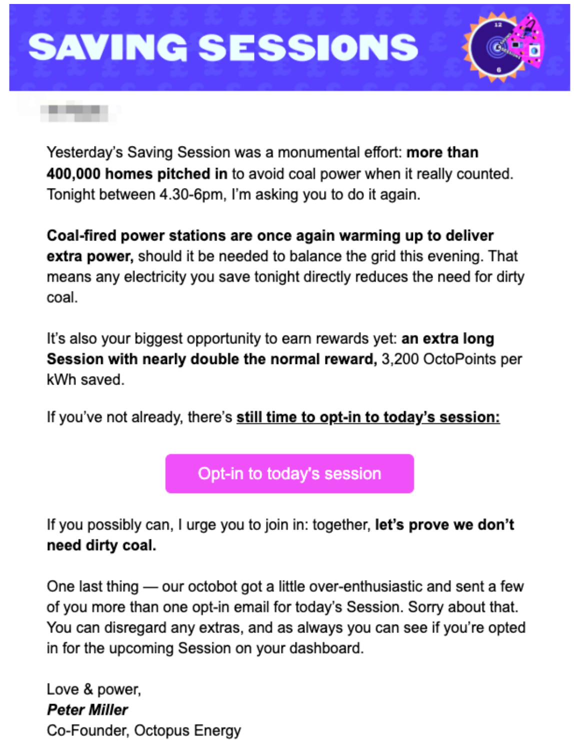
If not, there's **still time to opt-in to today's session:**

Opt-in

If you possibly can, I urge you to join in: together, **let's prove we don't need dirty coal.**

Love & power,
Greg Jackson
Founder, Octopus Energy

(**a**) Extra Notice for January 23.



SAVING SESSIONS

Yesterday's Saving Session was a monumental effort: **more than 400,000 homes pitched in** to avoid coal power when it really counted. Tonight between 4.30-6pm, I'm asking you to do it again.

Coal-fired power stations are once again warming up to deliver extra power, should it be needed to balance the grid this evening. That means any electricity you save tonight directly reduces the need for dirty coal.

It's also your biggest opportunity to earn rewards yet: **an extra long Session with nearly double the normal reward,** 3,200 OctoPoints per kWh saved.

If you've not already, there's **still time to opt-in to today's session:**

Opt-in to today's session

If you possibly can, I urge you to join in: together, **let's prove we don't need dirty coal.**

One last thing — our octobot got a little over-enthusiastic and sent a few of you more than one opt-in email for today's Session. Sorry about that. You can disregard any extras, and as always you can see if you're opted in for the upcoming Session on your dashboard.

Love & power,
Peter Miller
Co-Founder, Octopus Energy

(**b**) Extra Notice for January 24.

Note: Additional opt-in email notices sent to customers for the live DFS events which took place on January 23 and January 24, 2023. These sessions had a different remuneration per kWh of demand reduction — 2,700 OctoPoints in the former and 3,200 in the latter, equivalent to £3.375 and £4 per kWh respectively.

Communication with signed up customers worked differently for the two "live" Saving Sessions. Specifically, additional information was given to customers for the live events — where the email-based notices explained that the Saving Sessions on January 23 and 24 were especially important in terms of alleviating the need for Great Britain's national grid to harness electricity from coal and gas-based power stations. In addition to the higher-context environmental appeal (Bergquist et al., 2023), the two live Sessions featured higher remuneration. Specifically, the January 23 Session awarded customers 2,700 OctoPoints per kWh of demand reduction (£3.375/kWh) and the January 24 Session awarded customers 3,200 OctoPoints per kWh (£4/kWh).

2.4 Calculating Demand Reduction During the Demand Flexibility Service

2.4.1 NGESO's Prescribed Approach

Under the potential outcomes framework, a causal estimate of demand reduction is the difference between a customer's *actual* consumption during a given Saving Session and the consumption that *would have* occurred had this customer not signed up and not opted in to the event. However, we can only ever observe a customer's actual consumption. Thus, one must estimate their counterfactual consumption.[16]

[16] Potential outcomes are simply one's value for the response variable when simultaneously exposed to different experimental conditions. For a given individual i, their potential outcome under treatment (y_i^1) and their potential outcome under no treatment (y_i^0) cannot both be observed resulting in the "fundamental problem of causal inference". For instance, in a study of the effect of caffeine on hours of sleep, an individual i cannot receive and, at exactly the same point in time, not receive a cup of coffee prior to bedtime (Gelman et al. (2020, ch. 18)).

NGESO asked DFS providers to use a specific methodology to calculate each customer's counterfactual consumption, referred to as a customer's "baseline". When using NGESO's formula, DFS providers calculate demand reduction by subtracting a customer's "actual" (i.e., in-Session) consumption from their NGESO-methdology-derived "baseline" consumption. Thus, for a given customer i, a specific DFS event h and a given half-hour during this event t:

$$\text{Demand Reduction}_{iht} = \text{Consumption}_{\text{Baseline (NGESO Methodology)}_{iht}} - \text{Consumption}_{\text{Actual}_{iht}}, \tag{1}$$

where positive values indicate that, compared to their baseline, a customer used less electricity during a given portion of an energy-savings event.

Per NGESO guidance, customer baselines were estimated in accordance with *Balancing and Settlement Code (BSC) P376*, a legal framework governing how DFS providers ought to derive baseline consumption (Elexon BSC, 2023). Specifically, for each half-hour of a DFS event, a DFS provider was to calculate baseline consumption and demand reduction using the following three steps:

1. Calculate "unadjusted" baseline consumption by taking the unweighted average of consumption during the same half-hour of the day for the ten most recent weekdays.[17]

2. For each half-hour of a DFS event, subtract a customer's actual, in-event consumption from their baseline consumption *earlier on the same day* and then add this day-of difference to the "unadjusted" baseline to create an "adjusted" baseline consumption. For example, if the unadjusted baseline for a half-hour during a DFS event had been 0.4 kWh, an adjustment of 0.25 kWh would lead to an adjusted baseline during that half-hour of 0.65 kWh; whereas an adjustment of -0.25 kWh would lead to an adjusted baseline during that half-hour of 0.15 kWh.[18]

3. "Clip" demand reduction — i.e., code half-hours with negative demand reduction, that is when actual consumption was higher than "adjusted" baseline consumption (Equation (1)), as 0 kWh of demand reduction.[19]

DFS providers were not required to structure their remuneration of customers in the same manner. Indeed, they could have chosen any payment structure, such as an equal dividend to all DFS-participating customers based on average demand reduction. However, none did. Instead, all DFS providers structured customer payments as per-kWh demand reduction, echoing the payment structure they themselves received from NGESO. Furthermore, to calculate customer-level demand reduction, DFS providers all adopted the same formula NGESO used to calculate portfolio-level demand reduction.

The rationale behind NGESO's day-of adjustment (Step 2, above) was to give customers credit for days where energy consumption was structurally higher than the unadjusted baseline would suggest (e.g., due to cold weather) and to adjust downward when energy consumption was structurally lower than the unadjusted baseline would suggest (e.g., due to warm weather; (National Grid, 2023*e*)). However, anecdotal reports of customers increasing

[17] These weekday days must all fall within the last 60 days. If, for reasons such as faulty meter readings, there are no valid readings for any run of ten weekday days within the last 60 days, a DFS provider may use fewer working days to calculate the unadjusted baseline. However, DFS providers were instructed to exclude a customer entirely if they have fewer than five weekdays with valid readings in the last 60 days.

[18] Where the adjustment is negative and its absolute value is greater than the unadjusted baseline (e.g, an adjustment of -0.5 kWh against an unadjusted baseline of 0.4 kWh), the adjusted baseline is coded as 0 kWh rather than -0.1 kWh in order to avoid negative baselines which would be nonsensical. During half-hours with a 0 kWh baseline, demand reduction is not possible.

[19] For example, if a customer's "adjusted" baseline was 0.5 kWh, and their actual consumption was 0.8 kWh, their unclipped reduction would be -0.3 kWh. Clipping would transform this negative demand reduction into 0 kWh. Put formally, clipped demand reduction is, per Code P376, intentionally censored at the value of zero. Clipping is applied to each half-hour individually. Thus, if a customer's unclipped demand reduction was negative (e.g., -0.3 kWh) for the first half-hour of a 60-minute event, but positive (e.g., 0.2 kWh) for the second half-hour, half-hour-specific clipping would result in a total "clipped" demand reduction for the entire DFS event of 0.2 kWh.

their consumption as much as possible during the adjustment period in order to inflate their adjusted baseline became somewhat controversial during the DFS (Grimwood, 2023, National Grid, 2023*k*). NGESO has arranged for the Winter 2023-24 DFS baselining formula to *not* involve a day-of adjustment (National Grid, 2023*c*).

The rationale behind NGESO asking DFS providers to clip half-hourly demand reduction to a minimum value of zero (Step 3, above) was to avoid penalizing customers. Specifically, it was feared that penalties for negative demand reduction (i.e., when in-Session consumption was greater than adjusted baseline consumption) could harm consumers or cause risk-averse customers to decide not to participate in the DFS. Note that we observe large differences in demand reduction when using "clipped" versus "unclipped" measures. On average, across the 13 Saving Sessions delivered by Octopus Energy, clipped demand reduction among signed up customers was more than twice as high as unclipped demand reduction Figure AF.4, though the bias is smaller when the sample includes only customers who had signed up *and opted in*.

2.4.2 Methodology for Obtaining Our Counterfactual for the DFS and Estimation of Related Causal Effects

To estimate the causal impact of the DFS, we used difference-in-differences (DiD) to make three comparisons of special groups of customers:

- **Comparison 1**: A treatment group comprised of Octopus Energy customers who signed up to take part in DFS events before the first Saving Session on November 15, 2022 versus a counterfactual group comprised of Octopus Energy customers who never signed up to take part in DFS events (hereafter, *Signed Up Early* versus *Never Signed Up*).
- **Comparison 2**: A treatment group comprised of Octopus Energy customers who signed up to take part in DFS events before the first Saving Session versus a counterfactual group comprised of Octopus customers who signed up after the ninth Saving Session on January 30, 2023 (hereafter, *Signed Up Early* versus *Signed Up Late*; a robustness check for Comparison 1).
- **Comparison 3**: A treatment group comprised of Octopus Energy customers invited to sign up to take part in DFS events versus a counterfactual group comprised of newly-acquired smart-meter Bulb Energy customers who were unable to be invited to sign up (hereafter, *Octopus Customers* versus *Bulb Customers*).

We used our first two DiD strategies to examine the impact of signing-up to Saving Sessions on energy consumption *during Saving Sessions*, taking the difference between customers' in-Session consumption and their pre-treatment consumption before the first Saving Session. These two DiD strategies result in a causal estimand that is akin to an Intent-to-treat (ITT) effect. Put alternatively, we view actually opting into a Saving Session as our "treatment"; however, this treatment involves *non-compliance* because not all signed-up customers formally opt in to participate in one or more specific Saving Sessions(see Gelman et al. (2020, Ch. 21)). In this respect, signing up to Saving Sessions functions as a kind of "encouragment". And those who signed-up to Saving Sessions are merely eligible for treatment — where the ITT effect represents the demand reduction stemming from one's eligibility to opt into one or more the energy-saving events. We also estimated the local average treatment effect (LATE) using our first two DiD strategies by instrumenting the multiplicative interaction between sign-up and Saving Session opt-in.[20]. In contrast to the ITT, this allowed us to gauge the causal impact of actually participating in a Saving Session on in-Session consumption among those Octopus customers whose opt-in-related behavior could have been

[20]Formally, and following Gelman et al. (2020), the LATEs we estimate using our DiD designs are most appropriately labelled Compiler Average Causal Effects (CACEs). We use the more general "LATE" terminology here whereas we use the more specific "CACE" terminology below to draw a distinction between our field trial, which is analyzed through the lens of a randomised encouragement design, and our regression discontinuity design — the latter of which yields an estimand that is also a LATE.

altered by signing up to DFS events (cf. customers who signed up but who did not opt into a particular Saving Session).

Our third DiD strategy, a kind of natural experiment, was used to estimate an ITT effect summarizing the change in consumption stemming from eligibility to take part in the DFS itself. Specifically, we compared all Octopus Energy customers *invited* to take part in DFS events to all Bulb Energy customers who had a smart-meter and who would have been eligible to be invited to participate in DFS events had they been acquired by Octopus earlier in 2022.[21] This comparison is of substantial policy-relevance as Octopus Energy implemented a variant of the DFS (i.e., the 13 Saving Sessions) whereas Bulb Energy did not. *Thus, our comparison of all invite-eligible customers between the two suppliers clarifies the impact of supplier participation in the DFS* — recalling that NGESO recruited suppliers as opposed to individual households.[22]

"Treatment" for our third DiD strategy (i.e., actually opting into one or more specific Saving Sessions) involves *two* layers of non-compliance as: (a) not all Octopus customers encouraged to sign up to take part in DFS events via the aforementioned invitation will do so; and, as with our first two DiD designs, (b) not all customers signed up to DFS events (another form of encouragement) will opt in to participate in one or more Saving Sessions. Thus we also estimated two LATEs. First, we instrumented supplier "assignment" (i.e., whether a customer used Octopus as a service provider or Bulb as a service provider) to estimate the impact of *signing up* to take part in DFS events on in-Session consumption. Second, we again instrumented the supplier "assignment" to find the impact of *opting in* to participate in specific Saving Sessions on in-Session consumption.

Construction of Comparison Groups and the Handling of Staggering. For each of our three DiD designs, we constructed the comparison groups such that there is no "staggered" roll-out of the intervention. In doing so, we avoided problems highlighted by Baker et al. (2022) and Goodman-Bacon (2021) with traditional DiD designs concerning treatments that are not uniformly given across the study population.

The sample for our first DiD design consisted of most of the 1,384,400 customers who were sent emails inviting them to sign up to Saving Sessions (Figure AF.12). In total, 342,906 of these customers signed up "early", i.e., before the first Saving Session on November 15, 2022, versus 689,708 invited customers who never signed up, where the remaining customers signed up at various points throughout winter 2022/23. After exclusions, our sample consisted of 332,195 customers who signed up "early" and 654,062 who never signed up. We excluded from our analysis: (a) customers moving home or changing supplier between September 1, 2022 and March 31, 2023; (b) non-domestic customers; and (c) customers with missing or insufficient consumption data.[23]

[21] By "eligible", we mean Bulb customers who had a smart meter and who had not disallowed their supplier from using their smart meter to measure half-hourly consumption. Note that the majority of Octopus Energy customers with smart meters allow Octopus Energy to collect half-hourly consumption data.

[22] We stress that Bulb Energy and Octopus Energy had important similarities. Both began growing in terms of their share of the domestic energy supply market in 2018. Both provided customers a competitive "standard" tariff, avoiding situations where customers who finished fixed-price contracts would be moved onto a less competitive "standard" tariff — a practice known as "tease and squeeze" across the British energy sector. Both obtained most of their new customers through price comparison websites, while also utilizing very similar referral schemes wherein referrers earned £50 credit for securing a new customer and referees earned £50 credit for being referred. And both made all of their electricity tariffs "100% renewable" by default — where, in Great Britain, electricity suppliers can claim that a tariff is "100% renewable" if they procure sufficient Renewable Energy Guarantees of Origin to cover their customers' annual consumption (Ofgem, 2023*b*). For these reasons, Bulb customers are an unusually suitable comparison group for Octopus customers. And, normally, it would have been impossible to compare individual-level data from two suppliers in this manner. However, Octopus Energy's acquisition of Bulb Energy in December 2022 enabled this rare opportunity. As mentioned in the introduction, Octopus' acquisition of Bulb was just late enough to ensure that Bulb customers were not, for the most part, invited to Saving Sessions. Indeed, only 553 Bulb customers were invited. Some 22,000 who were not invited *did* sign up for Saving Sessions, and we regard this small minority of Bulb sign-ups as being akin to "always-takers" in our analyses.

[23] We classified customers as having "missing" consumption data if they had no consumption data for the entire period of October 1, 2022 to November 14, 2022 which we used as a "pre-treatment" (discussed below) period or if they had no consumption data for the days of the 13 Saving Sessions (i.e., the "post-treatment" period). We classified customers as having "insufficient" consumption data if their consumption was measured for one of these periods but not both.

The sample for our second DiD design (Figure AF.13) consisted of 331,992 customers who signed up "early", i.e., before November 15, 2022,[24] and 12,438 customers who signed up "late", i.e., between February 1, 2023 and the final Saving Session on March 23, 2023.

For our third DiD design, we used a sample of ≈1.1m Octopus customers (Figure AF.14): (a) who were sent emails by Octopus Energy requesting that they sign up for Saving Sessions; (b) had sufficient consumption data; and (c) were not excluded due to moving out of their home at some point during the winter 2022/23 Saving Sessions period, being a non-domestic customer, or switching to a different supplier during the winter 2022/23 period. Our Bulb Energy sample consisted of 197,307 smart-meter customers with sufficient consumption data. Note that we excluded a large number of Bulb Energy customers from our analysis due to missing smart-meter data. This missingness was the result of issues around database migration during Octopus' acquisition of Bulb. And these issues ultimately prevented us from being able to analyze the full set of Bulb smart-meter customers. In total, we dropped from our analysis 295,770 of Bulb's 493,077 smart-meter customers.[25]

Common ITT Effect for All 13 Saving Sessions. We obtained the ITT effect for all three of our difference-in-difference designs by following the recommendation of Bertrand et al. (2004) to collapse time series data into just two observations, on for each of two periods of study — that is, one pre-treatment (i.e., "baseline") period observation and one post-treatment period observation. Put alternatively, we fit models to a dataset composed of customer-period observations t, where each customer has just two observations such that $t \in \{1, 2\}$.[26]

To do this, for each customer, we simply averaged their half-hourly consumption (in kWh) across special half-hours during the pre-treatment period and during the post-treatment period.

To clarify, the pre-treatment period ($t = 1$) was constructed using all half-hours from 09:00 to 22:00, inclusive, Monday to Friday from October 1, 2022 to November 14, 2022.[27] We took this approach to ensure that our pre-treatment period reflected consumption during half-hours that were qualitatively similar to those during which the 13 Saving Sessions took place (Table 1) — where Sessions only occurred on weekdays and never occurred overnight. For each customer in our sample, we constructed their pre-treatment period observation using only data on consumption from October 1, 2022 to November 14, 2022 and eschewed using data further back in time (e.g., August and September) to avoid temporal inconsistencies. In particular, we wanted to avoid using data from months wherein the weather and the daylight hours were especially different from the weather and daylight hours during which Saving Sessions occurred (i.e, November 15, 2022 through March 23, 2023). Note well that our results are not sensitive to how we constructed the pre-treatment period (Section AI.5).

The post-treatment period ($t = 2$) was constructed using only the half-hours during which each Saving Session occurred after Saving Sessions began on November 15, 2022. Similarly to our construction of the observation for the pre-treatment period, we created the observation for the post-treatment period by averaging, for each customer, their half-hourly consumption during Saving Sessions (Table 1). Recall that Saving Sessions varied in length. Typically, however, Sessions lasted two half-hours. Longer Sessions occurred on December 12, 2022 (four half-hours; 17:00 to 19:00) and January 23, 2023 (three half-hours; 16:30 and 18:00).

Bearing this all in mind, we fit regression models to obtain two ITT effects stemming from: (a) eligibility to

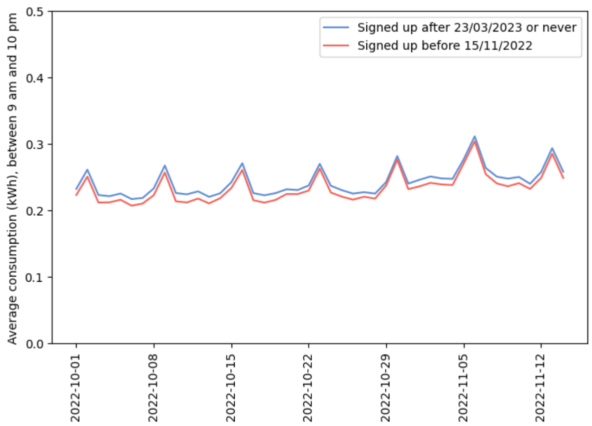
[24] The sample was slightly smaller in this DiD than in our first DiD because there were fewer Saving Sessions in the post-treatment period, which meant there were slightly more customers excluded for having insufficient consumption data in that period.

[25] As mentioned above, there were 531 Bulb customers who migrated to Octopus early enough to receive DFS invite emails along with Octopus Energy customers. For the purposes of our analysis, We classified these individuals as "regular" Octopus Energy customers in our first two DiDs. But, given their unusual categorization as "like" Bulb *and* "like" Octopus customers, we excluded them from our Octopus versus Bulb DiD.

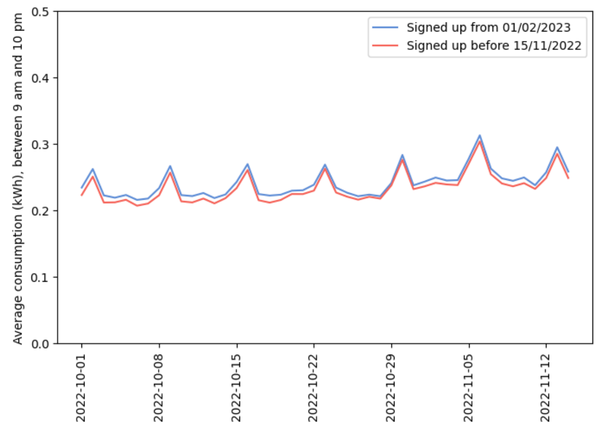
[26] This approach, which minimizes the number of customer-period observations that we used to fit our regression models, also allowed us to avoid excessive computational burden given the very large sample sizes involved in our analysis.

[27] In October, British customers were on British Summer Time. This means that the October half-hours are from 09:00 to 22:00 BST. Clocks in Great Britain "fell back" to Greenwich Mean Time on October 30, 2022.

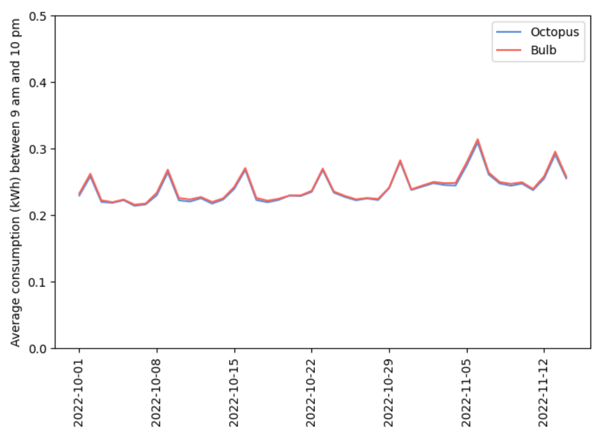
Figure 3: Parallel trends in average half-hourly electricity consumption (kWh) 09:00 to 22:00 from October 1, 2022, through November 14, 2022.



(**a**) DiD 1: Signed Up Early vs. Never Signed Up



(**b**) DiD 2: Signed Up Early vs. Signed Up Late



(**c**) DiD 3: Octopus Energy customers vs. Newly-Acquired Bulb Customers.

Note: Average half hourly consumption from 09:00 to 22:00, each day from October 1, 2022, through November 14, 2022. In the first two panels, Octopus Energy customers that signed up to take part in DFS events before November 15, 2022 (Red lines) are compared to customers who never signed up DFS events (Panel (a); Blue line) or to customers who signed up for DFS events on or after February 1, 2023 (Panel (b); Blue Line). In Panel (c), consumption among Octopus Energy customers who were invited to participate in DFS events (Blue line) are compared to Bulb customers with smart meters (Red line) over the same time period. In all three charts, peaks generally represent daytime consumption on Sundays, which, for domestic customers, tends to be higher than on other days.

opt into one or more saving sessions (DiD designs Signed up Early vs. Never and Signed up Early vs. Late); and (b) eligibility to sign up for the DFS itself (i.e., DiD design Octopus Customers vs. Bulb Customers). The two regression models have the following form:

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post-treat. Period},it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 (S_{\text{Post-treat. Period},it} \times T_{\text{Signed Up Early},it}) + \beta_4 H_{it} + \epsilon_{it} \tag{2a}$$

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post-treat. Period},it} + \beta_2 T_{\text{Octo Customer},it} + \beta_3 (S_{\text{Post-treat. Period},it} \times T_{\text{Octo Customer},it}) + \beta_4 H_{it} + \epsilon_{it}, \tag{2b}$$

where y_{it} is customer i's average half-hourly consumption in study period $t \in \{1, 2\}$ and $S_{\text{Post-treat. Period},it}$ is the binary indicator for treatment *period* — with $S_{\text{Post-treatment Period},it} = 0$ indicating the pre-treatment period (i.e., before the start of DFS events; October 1, 2022 to November 14, 2022; $t = 1$) and $S_{\text{Post-treatment Period},it} = 1$ indicating the post-treatment period (i.e., after the start of DFS events; November 15, 2022 through March 23, 2023; $t = 2$). Furthermore, $T_{\text{Signed Up Early},it}$ and $T_{\text{Octopus Customer},it}$ are the binary indicators for eligibility to receive treatment (i.e., encouragement to opt-in, as represented by signing up early, or encouragement to sign up for Saving Sessions, as represented by being invited to sign up). For our first two DiD designs (i.e., Signed Up Early vs. Never; Signed Up Early vs. Late; Equation (2a)), $T_{\text{Signed up Early},it} = 1$ for customers who signed up to take part in Saving Sessions by or on November 14, 2022 (i.e., the day before first Saving Session). And, for our third DiD design (i.e., Octopus Customers vs. Bulb Customers; Equation (2b)), $T_{\text{Octopus Customer},it} = 1$ for individuals who were a DFS-invited Octopus Energy smart-meter customer (rather than a Bulb Energy customer).

Accordingly, β_0, is the expected average half-hourly consumption in the *pre*-treatment period for *control-group customers* who never signed up (first DiD design), who signed up late (second DiD design), or who were a Bulb customer (third DiD design), β_1 is the expected difference in average half-hourly consumption between the post- and pre-treatment period, β_2 is the expected difference in average half-hourly consumption between the treatment and control groups owing to eligibility for Session opt-in (first and second DiD design) or the DFS itself (third DiD design), and β_3 — the coefficient for the multiplicative interactions ($S_{\text{Post-treatment Period},it} \times T_{\text{Signed Up Early},it}$) or ($S_{\text{Post-treatment Period},it} \times T_{\text{Octopus Customer},it}$) — is the expected difference in the slope coefficient for time period (i.e., β_1) between treatment groups (β_2). Thus, β_3 is the classic difference-in-differences (or double-difference) estimate involving just four group means (Gelman et al., 2020, p. 442). And, in the present case, β_3 is an ITT-type effect summarizing the causal impact of *eligibility* for treatment (i.e., Session opt-in or DFS participation) on average half-hourly consumption.[28]

Crucially, in using DiD, we necessarily make a (conditional) ignorability assumption in the following style:

$$d^0, d^1 \perp z \mid x, \tag{3}$$

where, as discussed by Gelman et al. (2020, p. 442-445), d^0, d^1 are, respectively and in the present scenario, the *potential change in average half-hourly consumption y_{it} between the post- and pre-treatment period* under no treatment and under treatment. Thus, by assuming conditional ignorability, we presumed that the distributions of potential changes are independent of treatment assignment among customers with the same value for the confounding variable x. Put simply, we assumed that the rate of change in y_{it} between the pre- and post-treatment period would be the same for the treatment and control groups in the absence of treatment (i.e., *parallel trends* in y_{it} over time). And, if this assumption holds, the post-pre difference in y_{it} for the control group (i.e., Never Signed Up Customers, Signed Up Late Customers, Bulb Customers), which is captured by β_1 in Equation (2) when $T_{it} = 0$, is a valid empirical counterfactual trend for individuals in the treatment group (i.e., Signed Up Early Octopus Customers and Octopus Customers in general) such that, were β_3 to be zero, β_2 would merely summarize the "jump" (i.e., difference in intercepts) between two parallel lines.

Note well that, for all three DiD designs and in the case of half-hourly consumption, we were highly unlikely to have fallen afoul of this particular assumption (see Figures 3a to 3c). Nevertheless, we still included in our linear predictors H_{it} — i.e., the average heating degree days (HDDs) in customer's i's region during period t. H_{it} is used

[28] To clarify the notion of "four means", and using Equation (2a) as an example while ignoring pre-treatment variables, note that the expected average half-hourly consumption for the control group in the pre-treatment period $\bar{y}_{S_{\text{Post-treatment Period},it}=0,\ T_{\text{Signed Up Early},it}=0} = (\beta_0)$, the expected average half-hourly consumption for the control group in the post-treatment period $\bar{y}_{S_{\text{Post-treatment Period},it}=1,\ T_{\text{Signed Up Early},it}=0} = (\beta_0+\beta_1)$, the expected average half-hourly consumption in the treatment group in the pre-treatment period $\bar{y}_{S_{\text{Post-treatment Period},it}=0,\ T_{\text{Signed Up Early},it}=1} = (\beta_0 + \beta_2)$, and the expected average half-hourly consumption for the treatment group in the post-treatment period $\bar{y}_{S_{\text{Post-treatment Period},it}=1,\ T_{\text{Signed Up Early},it}=1} = (\beta_0 + \beta_1 + \beta_2 + \beta_3)$. Thus, β_3 — i.e., the double difference of post-treatment period minus pre-treatment period and treatment group minus control group — is the result of $((\beta_0 + \beta_1 + \beta_2 + \beta_3)$ - $(\beta_0 + \beta_2))$ - $((\beta_0 + \beta_1)$ - $(\beta_0))$. And $((\beta_0 + \beta_1)$ - $(\beta_0))$ is the expected counterfactual change in average half-hourly consumption between the post-treatment and pre-treatment period that we assume to be valid for our treatment group by virtue of using a regression model (see Figure 3).

to adjust for the influence of outdoor temperature on domestic electricity consumption.[29] We do not anticipate that our results are vulnerable to the exclusion of other time-varying covariates.

Finally, the dataset we constructed to estimate the models summarized by Equations (2a) and (2b) is, fundamentally, hierarchical. That is, customer-period observations are nested *within* customers. Accordingly, we clustered standard errors at the level of individual customers using their meter point administration number (MPAN). Virtually all customers had properties with a single MPAN, where MPANs were the identifiers used for reconciliation of consumption and demand reduction for the purposes of Saving Sessions remuneration.

ITT Effect Specific to Each Saving Session. Equation (2) only provided us with common ITT effects for: (a) being eligible to opt into one or more saving sessions (Signed Up Early vs. Never Signed Up; Signed Up Early vs. Signed Up Late): and (b) being eligible to sign up for the DFS itself (i.e., Octopus Customers vs. Bulb Customers). Accordingly, we also considered heterogeneity in the three ITT effects across Winter 2022/23 by estimating variants specific to each of the 13 energy-saving events h.

To do this, we used the same pre-treatment period observation detailed above (i.e., $t = 1$) but modified how we constructed the singular post-treatment period observation (i.e., $t = 2$). Specifically, we constructed 13 post-treatment period observations wherein average half-hourly consumption y_{it} when $t = 2$ is calculated *only using recorded consumption for the half hours during which a Saving Session occurred* (Table 1). Thus, for each Saving Session $h \in \{1, ..., 13\}$, we separately fit regression models in the style of Equation (2) with the following general form:

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post-treat. Period } h,it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 (S_{\text{Post-treat. Period } h,it} \times T_{\text{Signed Up Early},it}) + \beta_4 H_{it} + \epsilon_{it} \quad \text{(4a)}$$

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post-treat. Period } h,it} + \beta_2 T_{\text{Octo Customer},it} + \beta_3 (S_{\text{Post-treat. Period } h,it} \times T_{\text{Octo Customer},it}) + \beta_4 H_{it} + \epsilon_{it}, \quad \text{(4b)}$$

where $S_{\text{Post-treatment Period } h,it}$, $T_{\text{Signed Up Early},it}$, and $T_{\text{Octopus Customer},it}$ are binary indicators as in Equation (2) save the Session restriction for $S_{\text{Post-treatment Period } h,it}$, β_3 is an ITT-type effect summarizing the causal impact of eligibility for treatment (i.e., Session-specific opt-in or participation in the DFS itself during a specific Saving Session) on average half-hourly consumption during a specific Saving Session h. We again clustered standard errors using MPANs.

Note that our second DiD design (i.e., Signed Up Early vs Signed Up Late) does not produce Session-specific estimates for Saving Sessions in February and March 2023. This is because customers' presence in the control group is, by definition, time limited. That is to say, to be in the control group for our second DiD design an individual must eventually sign up to take part in Saving Sessions, where we set the cut-off for sign-up as January 30, 2023 such that customers who signed up before this date (and customers who never signed up) were ineligible for inclusion in the control group for our second DiD design.

Common Local Average Treatment Effect (LATE) of Session Participation on in-Session Consumption. Recall that "treatment" for our three DiD designs — i.e., opting into a Saving Session — involves non-compliance. For our first and second DiD designs, not all customers who signed up early to DFS events ultimately go on to opt into one or more Saving Sessions — where those customers who do opt in may do so due to third factors that also govern consumption (i.e., opt-in is endogenous due to confounding). In this respect, being signed up to DFS events is a form of non-randomized encouragement received by those who signed up early but not by those who never signed

[29] Heating degree days (HDD) — i.e., the number of degrees that a day's average temperature is below some typical temperature — is a metric used to quantify the amount of energy demand for the purposes of heating a building in a 24-hour period. To calculate HDD, we drew hourly temperature data from various weather stations. In particular, for each of the 14 district networks in Great Britain, we gathered the average temperature of the weather stations in each district. We then created an HDD-like metric specific to a half hour t using $\max\{0, \frac{15.5℃ - \text{Avg. }℃_t}{48}\}$. To clarify, consider that, for instance, an average temperature above 15.5ºC during a half hour would count as 0 HDDs and that an average temperature of 15ºC during a half-hour would count as 0.5 ÷ 48 = 0.0104 HDDs.

up and only partially by those who signed up late. Yet, we would like to know the causal impact of actual Session participation on electricity use when aggregating across all 13 Saving Sessions.

Accordingly, we estimated the LATE of Session participation on in-Session consumption for our first and second DiD designs using a two-stage ordinary least-squares (2SLS) procedure (Gelman et al., 2020, Greene, 2019, Wooldridge, 2010) using the Python module "Linearmodels" (Sheppard et al., 2023). The first stage (i.e., the first model) was an OLS regression of an *aggregated version* of our treatment with imperfect compliance $P_{\text{Opt-in},it}$ — i.e., a continuous variable for the proportion of the 13 Saving Sessions a customer i opted into in the post-treatment period ($t = 2$) such that $P_{\text{Opt-in},it} := 0$ when $t = 1$ and $0 \leq P_{\text{Opt-in},it} \leq 1$ when $t = 2$ — conditional on our binary encouragement *in the post-treatment period* $t = 2$ as given by the multiplicative interaction ($S_{\text{Post-treat. Period},it} \times T_{\text{Signed Up Early},it}$), itself a binary indicator. The second stage (i.e., the second model) was an OLS regression of customer i's period-specific average half-hourly consumption y_{it} conditional on the predicted value of $P_{\text{Opt-in},it}$ from our first model, i.e., $\hat{P}_{\text{Opt-in},it}$.

Formally, the LATE for aggregate Session participation $P_{\text{Opt-in},it}$ for both our first and second DiD strategy, which we estimated using a joint procedure to correct the standard errors in the second stage (Gelman et al., 2020) while clustering standard errors using MPANs, were obtained using regression models with the following general form:

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post. Period},it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 \hat{P}_{\text{Opt-in},it} + \beta_4 H_{it} + \epsilon_{it,y} \tag{5a}$$

$$P_{\text{Opt-in},it} = \gamma_0 + \gamma_1 S_{\text{Post. Period},it} + \gamma_2 T_{\text{Signed Up Early},it} + \gamma_3 (S_{\text{Post. Period},it} \times T_{\text{Signed Up Early},it}) + \gamma_4 H_{it} + \epsilon_{it,P_{\text{Opt-in}}} \tag{5b}$$

$$P_{\text{Opt-in},it} = \begin{cases} 0 & \text{if } t = 1 \\ \text{Proportion of 13 Saving Sessions } i \text{ opted into.} & \text{if } t = 2, \end{cases}$$

where the linear predictor for $P_{\text{Opt-in},it}$ (i.e., Equation (5b)) and the linear predictor for y_{it} (i.e., Equation (5a)) are the first- and second-stage equations, $S_{\text{Post-treat. Period},it}$ and $T_{\text{Signed Up Early},it}$ are binary indicators, the latter of which indicates, depending on DiD design, customers who never signed up or customers who signed up late when $T_{\text{Signed Up Early},it} = 0$, $P_{\text{Opt-in},it}$ is the endogenous *continuous* treatment *for Octopus Energy customers who signed-up early and opted in to various numbers of Saving Sessions above zero*, $\hat{P}_{\text{Opt-in},it}$ is the predicted value for this continuous treatment, ($S_{\text{Post-treat. Period},it} \times T_{\text{Signed Up Early},it}$) is our non-random binary instrument/encouragement in the post-treatment period ($t = 2$), and H_{it} is the average heating degree days in customer's i's region during period t.

Given this set-up, γ_3 in Equation (5b) is the expected difference in the slope coefficient for time period between treatment groups and it captures the expected level of compliance in the post-treatment period for those in the treatment group in relation to all 13 Saving Sessions. And β_3 in Equation (5a) is the LATE representing the average causal effect of Session participation *when moving from participation in zero energy-saving events (i.e., $P_{\text{Opt-in},it} = 0$) to participation in all 13 energy-saving events (i.e., $P_{\text{Opt-in},it} = 1$)* on average half-hourly in-Session consumption amongst those individuals who's compliance rate in relation to all 13 Saving Sessions increases — *regardless of level* — due to early DFS participation.[30]

[30] As $P_{\text{Opt-in},it}$ is a continuous variable summarizing the proportion of the 13 Saving Sessions that a customer i opted into, the LATE we estimated using Equation (5) is, technically, *not* identified without additional, strong assumptions. To clarify, and following Gelman et al. (2020, p. 430), we have used a binary instrument — i.e., ($S_{\text{Post-treatment Period},it} \times T_{\text{Signed Up Early},it}$) — and a continuous treatment (i.e., $P_{\text{Opt-in},it}$) which is tantamount to scenario wherein one *uses a single binary instrument to identify the causal effect of multiple different treatments* — i.e., one (ordered) treatment for each level of the continuous variable (here, $\{(0 \div 13), (1 \div 13), ..., (13 \div 13)\}$) — such that there are more treatments than there are instruments, where each treatment may produce different causal effects (e.g., the causal impact of one's compliance with participating in two Saving Sessions as distinct from the causal impact of one's compliance with participating in ten Saving Sessions). Thus, we assume that our instrument is relevant to all 13 levels of our continuous treatment $\{(0 \div 13), (1 \div 13), ..., (13 \div 13)\}$, that our instrument induces groups of compliers who are similar in their unobserved characteristics across all levels of our continuous treatment, and that the causal effects of compliance among these latent sub-populations are "homogenous" and thus do not vary across levels of $P_{\text{Opt-in},it}$ in a manner whereby the β_3 is not an accurate summary of the LATEs for *all* complier groups.

Note that we include $S_{\text{Post-treatment Period},it}$ and $T_{\text{Signed Up Early},it}$ in Equations (5a) and (5b) to adhere to best practices for using multiplicative interactions and to reflect typical applications of 2SLS whereby the first- and second-stage linear predictors are identical save for instrument and the associated treatment-with-non-compliance.

LATE of Session Participation on Consumption During Each Saving Sessions. For our first and second DiD designs, we also obtained the LATE for each Saving Session individually using a setup (Equation (6)) that mixes those summarised in Equations (4) and (5). Note well that these models concern $P_{\text{Opt-in }h,it}$ — i.e., a binary treatment indicator for whether or not a customer i opted into a Saving Session h — as opposed to $P_{\text{Opt-in},it}$, the continuous treatment variable for the proportion of Saving Sessions that i opted into. Thus, in this instance, our first-stage models for $P_{\text{Opt-in }h,it}$ are linear probability models for a binary outcome and the issues discussed in Footnote 30 do not apply. Session-specific LATEs were estimated using 2SLS models with the following general form and clustered standard errors based on MPANs:

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post. Period }h,it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 \hat{P}_{\text{Opt-in }h,it} + \beta_4 H_{it} + \epsilon_{it,y} \tag{6a}$$

$$P_{\text{Opt-in }h,it} = \gamma_0 + \gamma_1 S_{\text{Post. Period }h,it} + \gamma_2 T_{\text{Signed Up Early},it} + \gamma_3 (S_{\text{Post. Period }h,it} \times T_{\text{Early},it}) + \gamma_4 H_{it} + \epsilon_{it,P_{\text{Opt-in }h}} \tag{6b}$$

$$P_{\text{Opt-in }h,it} = \begin{cases} 0 & \text{if } t = 1 \\ \text{Customer } i \text{ opted in to Saving Session } h? & \text{if } t = 2, \end{cases}$$

where $S_{\text{Post-treatment Period},it}$ and $T_{\text{Signed Up Early},it}$ are binary indicators, the latter of which indicates, depending on DiD design, customers who never signed up or customers who signed up late when $T_{\text{Signed Up Early},it} = 0$.

Two-Layered Non-Compliance and the LATE of DFS Sign-up and Session Participation When Comparing Octopus and Bulb Customers. Recall that "treatment" for our third DiD design involves two layers of non-compliance as: (a) not all Octopus customers invited (i.e., encouraged) to sign up to take part in DFS events did so; and (b) not all customers signed up to DFS events (itself a form of encouragement) opted in to participate in one or more Saving Sessions. Thus, customers who signed up and subsequently opted in may do so due to third factors that also influence consumption (i.e., both sign-up and opt-in are endogenous due to confounding). As with our first and second DiD designs, we would like to know the causal impact of actual Session participation, as well as the causal impact of actually signing up, on electricity usage when aggregating across all 13 Saving Sessions.

To manage non-compliance we used four models (Equations (7) to (9)) with clustered standard errors based on customers' MPANs. Specifically, we estimated the common LATE of DFS sign-up on in-Session consumption and the common LATE for Session participation on in-Session consumption for our third DID design using a two-stage least-squares (2SLS) regression models with the following general form — where the issues related to using a binary instrument for a continuous treatment discussed in Footnote 30 apply:

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post. Period},it} + \beta_2 T_{\text{Octo Cust.},it} + \beta_3 \hat{P}_{\text{Signed},it} + \beta_4 H_{it} + \epsilon_{it,y} \tag{7a}$$

$$P_{\text{Signed},it} = \gamma_0 + \gamma_1 S_{\text{Post. Period},it} + \gamma_2 T_{\text{Octo Cust.},it} + \gamma_3 (S_{\text{Post. Period},it} \times T_{\text{Octo Cust.},it}) + \gamma_4 H_{it} + \epsilon_{it,P_{\text{Signed}}} \tag{7b}$$

$$P_{\text{Signed},it} = \begin{cases} 0 & \text{if } t = 1 \\ \text{Proportion of 13 Saving Sessions remaining after } i \text{ signed up to DFS events.} & \text{if } t = 2 \end{cases}$$

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post. Period},it} + \beta_2 T_{\text{Octo Cust.},it} + \beta_3 \hat{P}_{\text{Opt-in},it} + \beta_4 H_{it} + \beta_5 + \epsilon_{it,y} \tag{8a}$$

$$P_{\text{Opt-in},it} = \gamma_0 + \gamma_1 S_{\text{Post. Period},it} + \gamma_2 T_{\text{Octo Cust.},it} + \gamma_3 (S_{\text{Post. Period},it} \times T_{\text{Octo Cust.},it}) + \gamma_4 H_{it} + \epsilon_{it,P_{\text{Opt-in}}} \tag{8b}$$

$$P_{\text{Opt-in},it} = \begin{cases} 0 & \text{if } t = 1 \\ \text{Proportion of 13 Saving Sessions } i \text{ opted into.} & \text{if } t = 2, \end{cases}$$

In both Equations (7) and (8), $S_{\text{Post-treatment Period},it}$ and $T_{\text{Octopus customers},it}$ are binary indicators for time and treatment group, where $T_{\text{Octopus customers},it} = 0$ indicating *Bulb* customers, $P_{\text{Signed},it}$ and $P_{\text{Opt-in},it}$ are *aggregated version* of our treatments with imperfect compliance, $\hat{P}_{\text{Signed},it}$ and $\hat{P}_{\text{Opt-in},it}$ are their predicted values, and the multiplicative interaction $(S_{\text{Post-treatment Period},it} \times T_{\text{Octopus Customer},it})$ is our binary encouragement *in the post-treatment period*.

Session-specific LATEs for sign-up and opt-in for our third DID design were estimated using 2SLS models with the following general form where, like Equation (6), the issues discussed in Footnote 30 do not apply:

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post. Period } h,it} + \beta_2 T_{\text{Octo Cust.},it} + \beta_3 \hat{P}_{\text{Signed } h,it} + \beta_4 H_{it} + \epsilon_{it,y} \tag{9a}$$

$$P_{\text{Signed } h,it} = \gamma_0 + \gamma_1 S_{\text{Post. Period } h,it} + \gamma_2 T_{\text{Octo Cust.},it} + \gamma_3 (S_{\text{Post. Period } h,it} \times T_{\text{Octo Cust.},it}) + \gamma_4 H_{it} + \epsilon_{it,P_{\text{Signed } h}} \tag{9b}$$

$$P_{\text{Signed } h,it} = \begin{cases} 0 & \text{if } t = 1 \\ \text{Customer } i \text{ signed up to DFS events before Saving Session } h\text{?} & \text{if } t = 2 \end{cases}$$

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post. Period } h,it} + \beta_2 T_{\text{Octo Cust.},it} + \beta_3 \hat{P}_{\text{Opt-in } h,it} + \beta_4 H_{it} + \epsilon_{it,y} \tag{10a}$$

$$P_{\text{Opt-in } h,it} = \gamma_0 + \gamma_1 S_{\text{Post. Period } h,it} + \gamma_2 T_{\text{Octo Cust.},it} + \gamma_3 (S_{\text{Post. Period } h,it} \times T_{\text{Octo Cust.},it}) + \gamma_4 H_{it} + \epsilon_{it,P_{\text{Opt-in } h}} \tag{10b}$$

$$P_{\text{Opt-in } h,it} = \begin{cases} 0 & \text{if } t = 1 \\ \text{Customer } i \text{ opted in to Saving Session } h\text{?} & \text{if } t = 2, \end{cases}$$

where $P_{\text{Signed } h,it}$ (Equation (9)) and $P_{\text{Opt-in } h,it}$ (Equation (10)) are binary indicators for our treatments with imperfect compliance, $\hat{P}_{\text{Signed } h,it}$ and $\hat{P}_{\text{Opt-in } h,it}$ are their predicted values, and the multiplicative interaction $(S_{\text{Post-treatment Period } h,it} \times T_{\text{Octopus Customer},it})$ is our binary encouragement in the post-treatment period.

Is Demand Displaced or Destroyed? A major point of interest for policymakers, grid operators, and energy retailers is whether reduced electricity consumption during DFS events influences demand during periods of time that are immediately adjacent to those half-hours within which an energy-saving events occurs (i.e., the half-hours immediately before and immediately after a Saving Session). Among those in the British energy sector, increased demand prior to or in the wake of an energy-saving events is taken as evidence of "demand displacement". In contrast, a lack of meaningful change in demand during the periods around an energy-saving event is taken as evidence of "demand destruction". *Decreased* demand prior to or in the wake of an energy-saving events is also taken as evidence of "demand destruction". This latter scenario can be viewed as a temporal spillover phenomenon whereby restricted use of electricity during an energy-saving event is associated with reduced consumption in the surrounding time periods.

Formally, demand destruction and demand displacement are statements about the *impact of reduced demand during an energy-saving event on temporally-adjacent electricity consumption*. This is in contrast to the impact of the event itself on consumption during adjacent time periods. However, in keeping with our other analyses, we used our first DiD strategy (i.e., Signed Up Early vs. Never Signed Up) to explore whether: (a) mere eligibility for participation in one or more Saving Sessions; and (b) actual Session participation (i.e., opt-in) are associated with temporally-adjacent electricity consumption in a manner that might be suggestive of displacement or destruction.

To do this, let us, for the purposes of this subsection, redefine the above set of period indices $t \in \{1, 2\}$ as $t \in \{1, 3\}$ whereby $t = 1$ continues to indicate our pre-treatment period (October 1, 2022 to November 14, 2022) but $t = 3$ now indicates the post-treatment period. Accordingly, to gauge the possibility of demand destruction or demand displacement, we pooled our data across all 13 Saving Sessions to construct two additional single-period observations in the vein of Bertrand et al. (2004) — i.e., a pre-Session period ($t = 2$) and a post-Session period ($t = 4$) — which temporally flank our original single-period post-treatment observations for our first DiD strategy discussed above such that now $t \in \{1, 2, 3, 4\}$. The pre-Session period ($t = 2$) was constructed by averaging each customer i's consumption across each pair of half-hours immediately before each Saving Session $h \in \{1, ..., 13\}$. The post-Session period ($t = 4$) was constructed by averaging each customer i's consumption across each pair of half-hours immediately after each Saving Session h.

Keeping this in mind, we wished to know the causal impact of eligibility for Session participation (i.e., the ITT effect) and actual Session participation (i.e., the LATE) on average half-hourly consumption when comparing our treatment and control group while also comparing $t = 1$ (the pre-treatment Period) to $t = 2$ (the pre-Session period) or $t = 1$ and $t = 4$ (the post-Session period).

Accordingly, to obtain the ITT effects, we fit two regression models (Equation (11)) that are variants of Equation (2) with the following form:

$$y_{it} = \beta_0 + \beta_1 S_{\text{Pre-Session Period},it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 (S_{\text{Pre-Session Period},it} \times T_{\text{Signed Up Early},it}) + \beta_4 H_{it} + \epsilon_{it} \quad \text{(11a)}$$

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post-Session Period},it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 (S_{\text{Post-Session Period},it} \times T_{\text{Signed Up Early},it}) + \beta_4 H_{it} + \epsilon_{it}, \quad \text{(11b)}$$

where $S_{\text{Pre-Session Period},it}$ and $S_{\text{Pre-Session Period},it}$ are binary indicators for time period.

The LATE for Session participation was estimated using two additional two-stage models (Equations (12) and (13)) that are variants of Equation (5) with the following form, noting that our aggregation across Sessions implicates issues related to using a binary instrument and a continuous treatment discussed in Footnote 30:

$$y_{it} = \beta_0 + \beta_1 S_{\text{Pre-Session Period},it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 \hat{P}_{\text{Opt-in},it} + \beta_4 H_{it} + \epsilon_{it,y} \quad \text{(12a)}$$

$$P_{\text{Opt-in},it} = \gamma_0 + \gamma_1 S_{\text{Pre-Session Period},it} + \gamma_2 T_{\text{Signed Up Early},it} + \gamma_3 (S_{\text{Pre-Sess. Period},it} \times T_{\text{Early},it}) + \gamma_4 H_{it} + \epsilon_{it,P_{\text{Opt-in}}} \quad \text{(12b)}$$

$$y_{it} = \beta_0 + \beta_1 S_{\text{Post-Session Period},it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 \hat{P}_{\text{Opt-in},it} + \beta_4 H_{it} + \epsilon_{it,y} \quad \text{(13a)}$$

$$P_{\text{Opt-in},it} = \gamma_0 + \gamma_1 S_{\text{Post-Session Period},it} + \gamma_2 T_{\text{Signed Up Early},it} + \gamma_3 (S_{\text{Post-Sess. Period},it} \times T_{\text{Early},it}) + \gamma_4 H_{it} + \epsilon_{it,P_{\text{Opt-in}}} \quad \text{(13b)}$$

$$P_{\text{Opt-in},it} = \begin{cases} 0 & \text{if } t = 1 \\ \text{Proportion of 13 Saving Sessions } i \text{ opted into.} & \text{if } t = 2 \vee 4 \text{ where } t \in \{1, 2, 3, 4\}, \end{cases}$$

Daily Demand. Using our first DiD strategy (i.e., Signed Up Early vs. Signed Up Never), we also considered the causal impact of eligibility for Session participation (i.e., the ITT effect) and actual Session participation (i.e., the LATE) on *daily* electricity consumption. Once again following Bertrand et al. (2004), we collapse time series data across multiple days (cf. multiple half-hours on specific days) into just two observations, one pre-treatment period observation ($t = 1$) and one post-treatment period observation ($t = 2$).

For the pre-treatment period in our day-level analysis, our response variable $y_{\text{Daily},it}$ is the average of customer i's daily consumption (i.e., the sum of consumption across all 48 half-hours) across all weekdays from October 1, 2022 to November 14, 2022. For the post-treatment period in our day-level analysis, $y_{\text{Daily},it}$ is an average of customer i's daily consumption across all 13 days that included a Saving Session (Table 1).

Accordingly, we estimated the ITT effect of eligibility for Session participation on average daily consumption using a regression model (Equation (14)) with MPAN-clustered standard errors and the following form:

$$y_{\text{Daily},it} = \beta_0 + \beta_1 S_{\text{Post-treat. Per.},it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 (S_{\text{Post-treat. Per.},it} \times T_{\text{Signed Up Early},it}) + \beta_4 H_{it} + \epsilon_{it} \tag{14}$$

And we estimated the LATE for Session participation using 2SLS and MPAN-based clustered standard errors, noting that our aggregation across Sessions implicates issues related to using a binary instrument and a continuous treatment discussed in Footnote 30:

$$y_{\text{Daily},it} = \beta_0 + \beta_1 S_{\text{Post-treat. Per.},it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 \hat{P}_{\text{Opt-in},it} + \beta_4 H_{it} + \epsilon_{it,y_{\text{Daily}}} \tag{15a}$$

$$P_{\text{Opt-in},it} = \gamma_0 + \gamma_1 S_{\text{Post-treat. Per.},it} + \gamma_2 T_{\text{Signed Up Early},it} + \gamma_3 (S_{\text{Post-treat. Per.},it} \times T_{\text{Early},it}) + \gamma_4 H_{it} + \epsilon_{it,P_{\text{Opt-in}}} \tag{15b}$$

$$P_{\text{Opt-in},it} = \begin{cases} 0 & \text{if } t = 1 \\ \text{Proportion of 13 Saving Sessions } i \text{ opted into.} & \text{if } t = 2 \end{cases}$$

For both Equations (14) and (15), y_{it} is customer i's average *daily* electricity consumption across the pre-treatment period ($t = 1$) or post-treatment period ($t = 2$), $S_{\text{Post-treatment Period},it}$ is the binary indicator for treatment period, $T_{\text{Signed Up Early},it}$ is the binary indicator for eligibility to receive treatment (i.e., encouragement to opt-in to specific Saving Sessions), and, somewhat differently to the other models, H_{it} is the average heating degree days *per day* in customer's i's region during period t.

Heterogeneity in Common ITT Effect Across Observable Customer Characteristics. Finally, we used our first DiD strategy (i.e., Signed Up Early vs. Never Signed Up) to consider heterogeneity in the causal impact of eligibility for Session participation (i.e., the ITT effect) when pooling across all 13 Saving Sessions. That is, we estimated our common ITT effect conditional on key traits of customers that we expect to modulate energy-related behaviours — namely, historical energy usage, the level of deprivation in one's geographic region, whether or not one has a smart-tariff, and the degree to which one's home is energy efficient. In this respect, the quantities we recover may be called "Conditional Average Treatment Effect" (CATEs), but note well that they are ITTs, not ATEs, where the ITTs are "diluted" due to non-compliance with treatment (i.e., Session participation).

To estimate conditional ITT effects, we extended Equation (2a) using *three-way multiplicative interactions*. And, to keep our model specifications tractable, we estimated the coefficients for the three-way interactions involving each customer characteristic x separately, where we also discretized continuous characteristics in order to incorporate them into our models using easily-interpretable binary indicators (i.e., "dummy" variables).

For instance, and focusing on Estimated Annual Consumption (kWh), which we discretize into the three categories "high" (EAC $\geq$ 2,900 kWh/year), "low" (EAC $<$ 2,900 kWh/year), and "Unknown/Missing"[31], we obtained conditional ITT effects using regression models with the general form:

$$\begin{aligned} y_{it} = \beta_0 + \beta_1 S_{\text{Post-treat. Period},it} + \beta_2 T_{\text{Signed Up Early},it} + \beta_3 (S_{\text{Post-treat. Period},it} \times T_{\text{Signed Up Early},it}) + \beta_4 H_{it} \\ + \beta_5 x_{\text{EAC = High},it} + \beta_6 (S_{\text{Post-treat. Period},it} \times x_{\text{EAC = High},it}) + \beta_7 (T_{\text{Signed Up Early},it} \times x_{\text{EAC = High},it}) \\ + \beta_8 x_{\text{EAC = Unknown},it} + \beta_9 (S_{\text{Post-treat. Period},it} \times x_{\text{EAC = Unknown},it}) + \beta_{10} (T_{\text{Signed Up Early},it} \times x_{\text{EAC = Unknown},it}) \\ + \beta_{11} (S_{\text{Post-treat. Period},it} \times T_{\text{Signed Up Early},it} \times x_{\text{EAC = High},it}) \\ + \beta_{12} (S_{\text{Post-treat. Period},it} \times T_{\text{Signed Up Early},it} \times x_{\text{EAC = Unknown},it}) + \epsilon_{it}, \end{aligned} \tag{16}$$

where $x_{\text{EAC = Low},it}$ is the reference category.

[31] *Estimated Annual Consumption (kWh)* is Octopus Energy's predicted customer consumption based on meter readings over the past year.

3 The consequences of participating in Saving Sessions: pre-posts and difference-in-differences

In this section, we show results from four approaches to estimate the treatment effect of the Savings Sessions from the Octopus Energy customers. First we used the approach that NGESO and DFS providers use to estimate demand reduction, a kind of pre-post effect within opted in households. Second, our first DiD approach compares differences between those who were invited to sign up to Savings Sessions and did so before the first Session versus those who were invited but never signed up. Third, our second DiD approach compares differences between those who were invited to sign up to Savings Sessions and did so before the first Session versus those that signed up late, in case there are unobservable characteristics associated with eventual sign-up that might threaten the validity of our first DiD. Fourth, we leveraged the natural experiment associated with Octopus Energy acquiring Bulb Energy.

3.1 Demand reduction as measured by NGESO methodology, a modified "pre-post" calculation

We examined the magnitude of demand reduction using NGESO's endorsed methodology to measure demand reduction, described in Section 2.4. According to this methodology, even groups that were not participating in Saving Sessions showed evidence of small demand reduction, as can be seen in Table 2. This pattern suggested that the standard "P376 baselining" methodology produced small biases overestimating demand reduction, at least during the 29 half-hours between November 15, 2022 and March 23, 2023 when the 13 Saving Sessions occurred.

In Table 2 and Figure AF.3, there were four groups of interest. First, there were those who signed up to Saving Sessions and opted in to the Session – we expect these customers' actual consumption to be lower than their baseline consumption on average, assuming that customers were on average successful in reducing their consumption. There were then three groups who did not opt in to the Session, for different reasons: 1) those who signed up but did not opt in to a Session; 2) those who had not signed up by the Session (or never signed up at all); and 3) Bulb customers who did not sign up to Saving Sessions.

Table 2: Demand reduction, measured using P376 methodology, by customer participation or non-participation.

Group	Proportion whose actual consumption $<$ baseline consumption	Average demand reduction (kWh per half-hour) during Saving Sessions	Customers $*$ half-hours
Signed up and opted in	74.4%	0.3052	8,878,988
Not signed up yet, or ever	54.6%	0.0236	33,664,210
Signed up, but not opted in	57.7%	0.0307	5,357,817
Bulb customer & never signed up	55.0%	0.0053	4,321,299

Note: We examined demand reduction according to the methodology prescribed by NGESO, depending on whether customers had opted in to a Saving Session. Here, we show this reduction for: customers who signed up and opted in to a Session, customers who signed up but did not opt in, invited Octopus Energy customers who never signed up, and Bulb customers who did not sign up. Approximately 55% of the customers in the latter three non-participating groups showed a small demand reduction; on average, their demand reduction was between 2% and 11% of the average demand reduction of customers who signed-up and opted-in.

With this said, as Table 2 shows, demand reduction – as measured by the "unclipped" version of the standard NGESO methodology – was much higher for customers who had signed up and opted in (0.305 kWh) than for the three sets of customers who were not participating (0.005 - 0.031 kWh demand reduction).

3.2 Demand reduction as measured by difference-in-differences approaches

Section 3.1 compared unclipped demand reduction based on P376 baselines between customers who participated in Sessions and those who did not. As discussed, unclipped demand reduction is a modified pre-post calculation. Comparing pre-posts between groups is, effectively, a series of informal DiDs. In this section, we conducted more formal DiD analyses to identify causal estimates of demand reduction from customers participating in Saving Sessions, during the Sessions. In all three of our DiDs, the key assumptions are parallel trends and no anticipation effects. We examined pre trends visually in Section AI.1. We conducted formal tests of parallel pre-treatment (pre-Saving-Session) trends in Section AI.4.

3.2.1 Impacts of participation on consumption during Saving Sessions

The ITT effects of Saving Sessions from our three DiDs differed, but in ways that made sense given their differing samples. In Figure 4 and Table AT.4 we show how the estimates differ by DiD strategy. As discussed in Section 2.4.2, the ITT for the first two DiD strategies is similar in interpretation to the first LATE (of sign-up) for the Octopus versus Bulb DiD. The LATE on opt-in can be interpreted similarly across the three estimation strategies. However, the Signed Up Early versus Late DiD had a smaller sample of Saving Sessions – only the first nine, rather than all 13.

Using our Octopus versus Bulb DiD, we found that simply inviting customers to sign up to Saving Sessions is associated with a 9.43% ($\pm$ 0.26% 95% confidence interval) reduction in consumption during Saving Sessions.[32] Using all three of our DiDs, we found that signing up to participate in DFS events reduced demand by $\approx$25% during Saving Sessions.[33] Additionally, we found that "opting in" to participate in Saving Sessions reduced demand by $\approx$40% during the campaign.[34]

In kWh terms, we found that signing up to Sessions caused a reduction during Sessions of $\approx$0.09 to 0.1 kWh per half-hour and opting in a reduction of $\approx$0.14 to 0.17kWh per half-hour (the ranges represent the slight differences in the results from our three DiDs; see Table AT.4). Customers not participating in Saving Sessions consumed $\approx$0.4 kWh per half-hour during the half hours when Saving Sessions occurred. To put these figures in perspective, the $\approx$0.4 kWh per half-hour consumption in the control groups was less than a dishwasher or clothes washer uses (0-1 kWh per half-hour) and much less than an oven or dryer uses (0.5-3 kWh per half-hour). Even a 60 watt bulb (0.03 kWh per half-hour, assuming it is not an LED) can be a meaningful source of reduction if turned off for an hour, though the bulb's consumption is much lower if it is an LED. In any given half-hour, these appliances consume more electricity on their own; but, of course, they are not on all day. Still, postponing or avoiding using them *can* have a large impact on half-hourly consumption.

In Figure 5, Figure 6a, and Figure 6b, we show the impact for each Saving Session, from the series of regressions where each Saving Session is its own post-treatment period, by the effect we're measuring – of being invited, signing up, or opting in. In these figures, we show the difference-in-differences impact as the percent of the "control" group's half-hourly consumption during each Saving Session. [35]

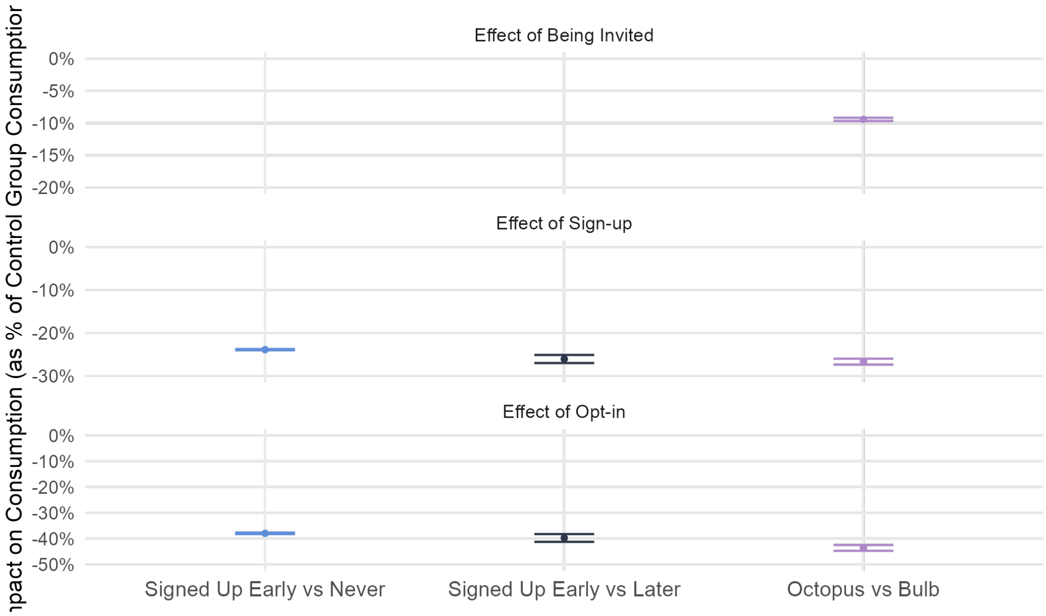
[32] Approximately half of Octopus Energy customers had smart meters, so the effect of Octopus Energy participating in DFS was approximately half of the 10% effect of invitation, i.e., a 5% demand reduction across Octopus Energy's smart- *and* traditional-meter customer base.

[33] Specifically: -23.88% ($\pm$0.13%) in the Signed Up Early versus Never DiD, -26.05% ($\pm$0.94%) in the Signed Up Early versus Late DiD, and 26.67% ($\pm$0.69%) in the Octopus versus Bulb DiD.

[34] Specifically: -37.94% ($\pm$0.29%) in the Signed Up Early versus Never DiD, -39.74% ($\pm$1.50%) in the Signed Up Early versus Late DiD, and -43.59% ($\pm$1.14%) in the Octopus versus Bulb DiD.

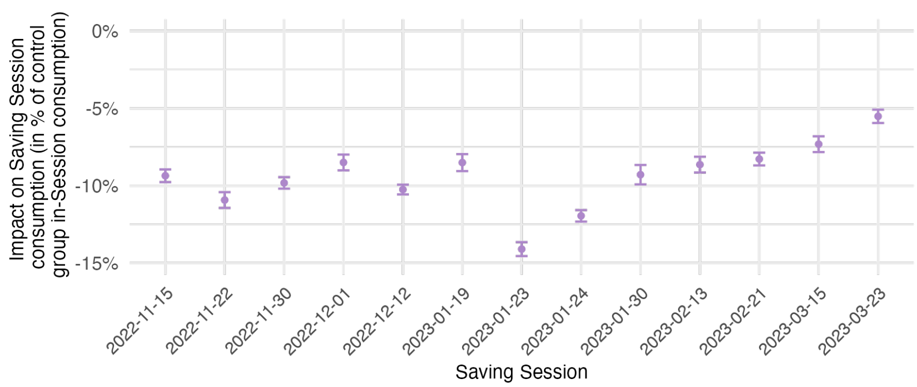
[35] In the appendix, we also show these impacts in kWh (the direct outcome measure in our DiD regressions): Tables AT.1 to AT.3, Figure AF.1, and Figure AF.2.

Figure 4: Difference-in-differences results.



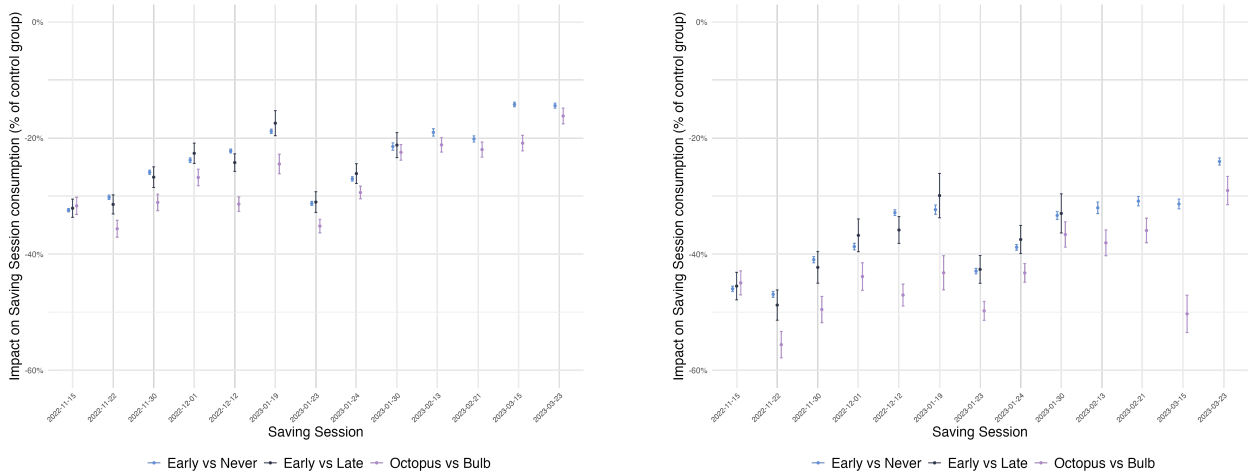
Note: The results of the three DiDs with 1) Customers who signed up before the 1st Saving Session versus customers who never signed up, 2) Customers who signed up before the 1st Saving Session versus customers who signed up after the 9th Saving Session and 3) Octopus customers invited to sign up versus smart-meter Bulb customers (who were not invited to participate in the DFS). We show point estimates and confidence intervals as a percent of each DiD's control group average half-hourly consumption during Saving Sessions. For our first two DiDs – 1) Signed Up Early versus Never, and 2) Signed Up Early versus Late – the ITT effect is the effect of sign-up. For the Octopus versus Bulb DiD, the ITT effect is the effect of being invited, and we estimate the LATE of sign-up. For the effect of opt-in, we use the LATE on opt-in from all three DiDs.

Figure 5: Difference-in-differences results: impact of being *invited to sign up*.



Note: Coefficient (and 95% confidence intervals, whiskers) on the difference-in-differences in our Octopus versus Bulb DiD for each of 13 regressions, where the post-treatment period in each regression was customers' half-hourly consumption during each of the 13 Saving Sessions. We interpreted this coefficient as the causal impact of being *invited to sign up* to Saving Sessions.

Figure 6: Difference-in-differences results.



(**a**) Impact of signing up.

(**b**) Impact of opting in.

]

Note: Coefficient (and 95% confidence intervals, whiskers) on the difference-in-differences in our three DiDs for each of 13 regressions, where the post-treatment period in each regression is customers' half-hourly consumption during each of the 13 Saving Sessions. In the Octopus versus Bulb DiD (left), the coefficient was on the local average treatment effect (LATE) of sign-up, a variable equal to 1 if a customer had signed up to Saving Sessions by that Session, else 0. We interpreted these coefficients as the causal impacts of being **signed up** to Saving Sessions by the date of the Session. In each DiD (right), the coefficient was on the local average treatment effect (LATE) of opt-in, a variable equal to 1 if a customer opted in to the Session, else 0. We interpreted these coefficients as the causal impacts of *opting in* to Saving Sessions on the date of the Session.

In examining demand reduction by event, we saw suggestive evidence that the treatment effect declined over the course of the DFS season, and that it seemed to rise back to initial levels on the "live" events on January 23 and 24, 2023, which involved higher incentives, more press, and extra reminders. The pattern of declining treatment effects over the winter cannot solely be attributed to *new* sign-ups reducing their demand by less than *early* sign-ups, which could attenuate the overall average demand reduction in the Octopus versus Bulb DiD. After all, we saw this same pattern even in the first two DiD strategies where the treatment group was always early sign-ups. Thus, the effect would have to be at least partially due to declining performance per individual signed up household. However, with only 13 Saving Sessions, this evidence on a potential declining level of demand reduction should be treated with caution.

Importantly, we compared our measures of demand reduction from each of our three DiDs to the estimates that NGESO's prescribed "P376" methodology would have calculated for the same samples in each analysis. The "unclipped" version of the P376 methodology, applied across each DiD's full treatment group (regardless of opt-in), approximated our DiD results well – 2.4 and 3.4% higher than our first two DiDs' estimates and 13.7% lower than our Octopus versus Bulb DiD estimate. As discussed in Section 2.4.1, the official version of the P376 methodology is the "clipped" version, and this resulted in large over-estimations – 70-80% in the first two DiDs and 264% in the third. However, the first two DiDs' signed up groups contained a mix of customers who opted in and did not opt in, and the Octopus versus Bulb DiD contains a mix of Octopus customers who signed up and did not sign up. Most of the upward bias came from these customers, whose unclipped demand reduction is near-zero but whose clipped demand reduction can be substantial, as negative baselining errors do not cancel out positive ones. Official demand reduction – as reported to NGESO from DFS providers such as Octopus Energy – was the clipped demand reduction of *opted in* customers only. This sample selection substantially reduced the bias from the official reported statistics. As we see in Table AI.5 and discuss further in Section 5, our DiD estimates were on average actually *higher* than demand reduction estimated by *unclipped* P376 among opt-ins, though still lower than the demand reduction estimated by *clipped* P376 among opt-ins.

Finally, note that results obtained using our first and third DiD strategies are not artifacts of what some readers may regard as an unusual combination of difference-in-differences with instrumental variables estimations (cf. "instrumented difference-in-differences"). Indeed, in Section AI.3, we show that are results are virtually unchanged when comparing LATEs obtained using our binary instruments (i.e., the multiplicative interactions between treatment group and treatment period) and 2SLS regression to Wald estimates derived "by hand" (Gelman et al., 2020, p. 426) — i.e., ratios that are the result of dividing our ITT effects for our first and third DiD designs from Equations (2) and (4) by the overall Saving Session opt-in rate and the Session-specific opt-in rates (i.e., the compliance rate). Our approach yields nearly identical figures for: (a) the common and Session-specific LATEs of Session *opt-in* on consumption obtained using our first DiD design (Table AI.1); (b) the common and Session-specific LATEs of DFS *sign-up* on consumption obtained using our third DiD design (Table AI.2); and (c) the common and Session-specific LATEs of Session *opt-in* on consumption obtained using our third DiD Design (Table AI.3).

3.2.2 Impacts of participation on consumption just before and just after Saving Sessions

We also examined the impact of Saving Sessions on consumption in the hour just before and just after Saving Sessions. As shown in Tables 3 and 4, we found evidence of small but meaningful "spillover" of reduction into neighboring half-hours.[36] In other words, we see evidence of demand destruction.

Table 3: Impact of Saving Sessions on hour just before Session (kWh per half-hour).

DiD approach	ITT	Average consumption (kwh per half-hour) during SS	ITT in %
Signed Up Early versus Never	-0.0079 (0.0005)	0.324	-2.44%
Signed Up Early versus Late	-0.0067 (0.0018)	0.312	-2.15%
Octopus versus Bulb	-0.0034 (0.0003)	0.326	-1.04%

Note: Coefficient (and standard errors, in parentheses) on the difference-in-differences in our three DiDs, where the post-treatment period in each regression was customers' half-hourly consumption during the hour before Saving Sessions. For our first two DiDs – 1) Signed Up Early versus Never, and 2) Signed Up Early versus Late – we interpreted these coefficients as the impact of *signing up* to Saving Sessions on consumption in the hour just before the Saving Session. For the Octopus versus Bulb DiD, we interpreted this coefficient as the impact of being *invited to sign up* to Saving Sessions on consumption in the hour just before the Saving Session.

To investigate this question further, we used this same method to go farther back and forward in time to produce Figure 7. There are some potential biases induced by this method in the context of our two-period DiD set-up as we go too far backward or forward in time – i.e., into overnight half-hours. The pre-treatment period in our DiDs is the average half-hourly consumption between 09:00 and 22:00 from weekdays in October and the first half of November. This consumption overlaps our post-treatment period, whether in our analyses of in-Sessions consumption or of consumption just before and just after Sessions. We also know that our treatment and control groups from each DiD have similar consumption profiles during those hours of the day. As we move far enough back or forward in time in the definition of our post-treatment period, however, this concordance between pre- and post-treatment periods breaks down, causing a potential violation of parallel trends. With this in mind, we avoided examining the impact of Saving Sessions on overnight consumption. In order to compare similar half-hours as we went farther back and forward in time, we constructed a new sample of Saving Sessions: those that started at 17:00 or 17:30 and

[36] In this case, the results differ by DiD strategy more than one might expect. The first two DiD strategies' ITT estimates accorded with each other, but the Octopus versus Bulb difference-in-difference produces overall higher impact estimates for the hour "just after" Saving Sessions. The LATE on sign-up for the Octopus versus Bulb DiD, for example, indicates a reduction of 0.0355 kWh per half-hour in the hour just after, in contrast to effects of approximately 0.01 kWh in the first two approaches.

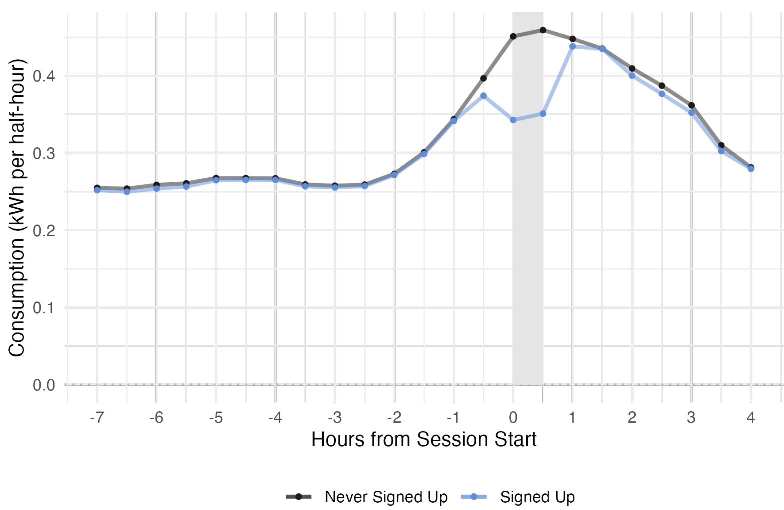
finished at 18:00 or 18:30.[37] We then implemented our DiD separately for each half-hour in the seven hours *before* and the 4 hours *after* these Sessions began (thus the first hour represents the two half-hours *during* the Session).

Table 4: Impact of Saving Sessions on hour just after Session (kWh per half-hour).

DiD approach	ITT	Average consumption (kwh per half-hour) during SS	ITT in %
Signed Up Early versus Never	-0.0089 (0.0005)	0.363	-2.45%
Signed Up Early versus Late	-0.0114 (0.0018)	0.367	-3.10%
Octopus versus Bulb	-0.0125 (0.0003)	0.373	-3.35%

Note: Coefficient (and standard errors, in parentheses) on difference-in-differences in our three DiDs, where the post-treatment period in each regression was customers' half-hourly consumption during the hour after Saving Sessions. For our first two DiDs – 1) Signed Up Early versus Never, and 2) Signed Up Early versus Late – we interpreted these coefficients as the impact of *signing up* to Saving Sessions on consumption in the hour just after the Saving Session. For the Octopus versus Bulb DiD, we interpreted this coefficient as the impact of being *invited to sign up* to Saving Sessions on consumption in the hour just after the Saving Session.

Figure 7: Event study of the six Saving Sessions that spanned 17:00 to 18:00 or 17:30 to 18:30.



Note: We show the coefficient on the DiD from a series of regressions using the Signed Up Early versus Never DiD sample. We interpret this coefficient as the demand reduction during the relevant half-hours caused by signing up to Saving Sessions (an ITT effect diluted by incomplete opt-in). The times "0" and "0.5" on the x-axis are from a regression where the post-treatment period is half-hourly consumption during Saving Sessions – two half-hours per Session. All other points on the x-axis are from regressions where the post-treatment period is a single half-hour – X hours before the Session started, if before, or X hours after the Session finished, if after. The values of the Never Signed Up group are their actual consumption during each half-hour during the days of the six Saving Sessions that panned 17:00 to 18:00 or 17:30 to 18:30. The values of the Sign Up (early) group are the sum of the Never Signed Up group consumption and the beta on the DiD interaction of the post-treatment period and the Sign Up indicator. We do not show the 95% confidence interval around each point, as the interval is too narrow on this graph to see clearly. In the appendix, we show another version of this graph (Figure AF.15) where the confidence intervals are visible.

We found evidence of substantial demand reduction during the Saving Session, and (as discussed immediately above) a much smaller but still meaningful demand reduction just before and just after. There may still be demand displacement to the half-hours overnight or the day before or after a Saving Session, but we did not believe our DiD approaches were robust to detecting such displacement, as opposed to potentially small but meaningful time-varying differences unrelated to Saving Sessions in the groups' overnight consumption.

[37] This left us with six sessions: Sessions 1, 2, 3, 4, 7, 10, and 11.

Finally, we looked at Saving Sessions' impact on *daily* consumption in Table 5. This was theoretically a direct way to answer whether Sessions caused demand displacement or destruction. Our goal was to examine whether Saving Sessions were associated with lower daily consumption above and beyond the direct impact during the Session, indicating demand destruction plus spillover destruction; lower daily consumption of similar magnitude to the impact during the Session, suggesting a simple demand destruction story; no change in daily consumption, suggesting simple demand displacement; or higher daily consumption, suggesting demand displacement and some extra creation. However, our daily consumption regression was less precise than our analyses of half-hourly consumption. While we saw a null result in our three DiDs (examining the ITT effect, diluted by incomplete opt-in and, for the Octopus versus Bulb DiD, sign-up), the 95% confidence interval is consistent with any of the four stories. The point estimate and therefore the weight of the confidence interval is consistent with substantial demand destruction – greater than the impact during the 2-4 Saving Session half-hours we identify in our primary analyses above of ≈0.2 kWh / Saving Session[38] – but the effect was too noisy to enable firm conclusions about effect magnitude or even direction.

Table 5: Impact of Saving Sessions on the day of Session (kWh per day).

	ITT	95% CI	Mean kwh per day on SS days, among "control" group	ITT as % of mean
Signed Up versus Never	-0.3769 (0.205)	[-0.78, 0.026]	21.298	-1.77%
Octopus versus Bulb	-0.4169 (0.227)	[-0.861, 0.027]	19.3635	-2.15%

Note: Coefficient (and standard errors, in parentheses) on difference-in-differences of our Signed Up Early versus Never and Octopus versus Bulb DiD (ITT effects, diluted by incomplete opt-in, and in the latter DiD incomplete sign-up). The post-treatment period in each regression is customers' average daily consumption on the days of Saving Sessions. The pre-treatment period is daily consumption on weekdays in October 2022 and the first two weeks of November 2022 (before the first Session on November 15, 2022).

3.2.3 Conditional Average Treatment Effects

We examined how the DiD estimates vary by observable customer characteristics, identifying Conditional Average Treatment Effects (CATEs) using Equation (16). As we see in Section AT.3, treatment effects were higher for customers from lower-deprivation postcodes, those with higher estimated annual consumption, and those on smart tariffs. However, even the groups with lower treatment effects still showed large, economically meaningful effects of participating in Saving Sessions. For example, looking at the CATEs in the first difference-in-differences strategy (Signed Up Early versus Never), the change in demand for customers on smart tariffs was -0.1395 kWh (25.0% of the during-Session half-hourly consumption of non-signed-up customers on smart tariffs). This extra 0.0514 kWh is a meaningful and and statistically significant difference from customers not on smart tariffs. However, the -0.0881 kWh demand change among customers not on smart tariffs (representing 19.8% of the during-Session half-hourly consumption of non-signed-up customers on non-smart tariffs) was still substantial and indeed close to the overall Average Treatment Effect of -0.0897 kWh. Similarly, the CATE for customers in very low deprivation postcodes was -0.1059 kWh (-22.7% of during-Session consumption of non-signed-up customers from very low deprivation postcodes), while the CATE for customers in very high deprivation postcodes was -0.0644 kWh (-17.2% of during-Session consumption of non-signed-up customers from very high deprivation postcodes). This difference was meaningful and statistically significant, but the demand reduction among customers in very high deprivation postcodes was still substantial. Finally, there was some heterogeneity with respect to customer EPC

[38] We refer here to the estimated impact of sign-up of 0.09kWh per half-hour, from our ITT derived in our Signed Up Early versus Never DiD; and our LATE of 0.1 kWh per half-hour from our LATE on sign-up from our Octopus versus Bulb DiD. Eleven of the 13 Saving Sessions had two half-hours, while one had three and one had four.

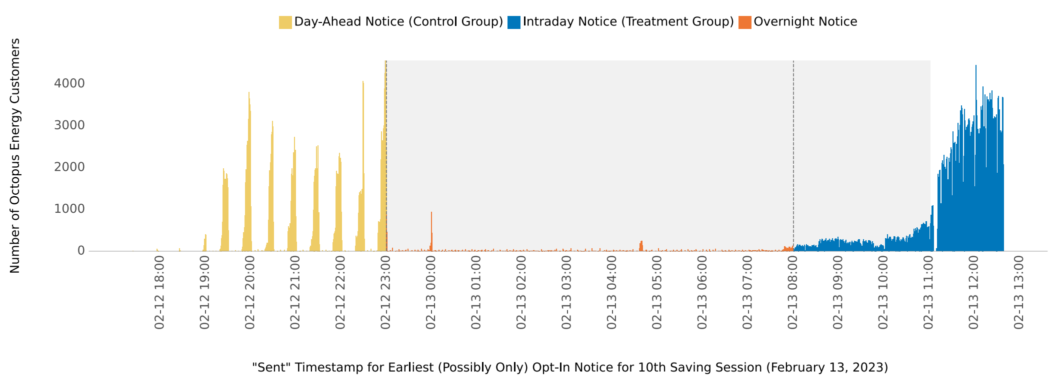
letter grade, but the effects were not monotonic; lower-grade homes showed higher demand reduction than homes with grades B, C, and D, but so do A-grade homes.

3.2.4 Mechanisms of Energy Reduction During Saving Sessions

As we describe fully in Section AI.2, we invited 55,000 randomly-chosen signed-up customers to answer questions about their experience of Saving Sessions. In total, 5,751 customers responded. To better understand the behavioral mechanisms of energy reduction, we asked survey respondents who opted into Saving Sessions "What best describes how you participated?" and provided them with a series of non-mutually-exclusive responses. Customers could "tick" agreement with as many of the responses as they wished. 75% of respondents indicated that they engaged in manual demand shifting in that they "manually switched off appliances during the Session and used them at other times". A much smaller group (i.e,. 22% of respondents) indicated that they incorporated *scheduled* demand shifting, agreeing to either (or both) of the response options "scheduled my appliances (like the tumble dryer) to come on before the Session" and/or "scheduled my appliances (like the tumble dryer) to come on after the Session". Note that we observed that these proportions did not vary much by subgroup: manual methods of demand shifting were more frequent then scheduled techniques amongst all sub-populations we examined. However, survey respondents on smart tariffs and those with higher estimated annual electricity consumption did *more* scheduling of appliances (Figures AF.20 and AF.21). We observe little difference on these measures in relation to customers' postcode-level deprivation (Figure AF.22).

4 Optimal Peak-Pricing Campaign Design

Figure 8: Distribution of the time at which an Octopus Energy customer was sent their first (possibly only) opt-in notice for the Saving Session on February 13, 2023.



Note: Shaded region denotes the window of time on which our analysis of energy consumption is concentrated and corresponds to roughly 22:57 on the 12th to 23:01 on the 13th (see Section 4.2 on "bandwidth"). For our analysis of agreement to participate in the 10th Saving Session we used a window of time corresponding to roughly 22:29 on the 12th and 10:52 on the 13th.

We now turn to our investigation of the design of peak-pricing campaigns, where we wish to know how transmission systems operators and utility firms might craft these initiatives to maximize their impact on energy behavior. Here we focused in particular on the broad timing (i.e., When?) of Octopus Energy's appeals to its customers to flexibly use electricity and the general channel through which these appeals were made (i.e., their "type"; e.g., email versus SMS). And we probed the causal impact of these characteristics on: (a) levels of in-Session consumption; and (b) the probability of Session participation.

4.1 Early Versus Late Timing of Saving Session Notices

First, we considered the causal impact of receiving an intraday opt-in notice as opposed to a day-ahead notice. The manner in which opt-in notices were sent for the Saving Session on February 13 (17:30 to 18:30) was not random. However, the unexpectedly-delayed time-ordered dispersal of notices (Figure 8), which we discussed in Section AI.8, was amenable to a regression discontinuity design (RDD).

Regression discontinuity (RD) is a quasi-experimental method used to analyze observational data where the mechanism by which a treatment, policy, or exposure was assigned (i.e., administered) is entirely known but there is *no* randomization. Specifically, given some "assignment" variable A (here, an Octopus Energy customer's account ID) used to administer some treatment z (here, intraday opt-in notice) to individuals $i \in N$ (here, all DFS-participating Octopus Energy customers as of February 12, 2023), RD is used to compare individuals whose values for the assignment variable A_i fall "just above and just below" a predetermined cut-off C. With some additional assumptions, and if individuals just above the cut-off (Group 1) and just below the cut-off (Group 2) are similar, a causal comparison can be made between the two groups with respect to an outcome of interest (here, in-Session consumption and Session participation).

We used a *temporal* cut-off C_{Time} equal to 08:00 on February 13, 2023 — where Octopus Energy customers sent opt-in notices at or after this time were assigned to our treatment group (i.e., receipt of an intraday notice as opposed to a day-ahead notice). Because our cutoff for treatment was temporal and not a specific account ID, we had to map our cutoff to an integer value reflective of the scale and the ordering of Octopus Energy customers' account IDs. We did this by selecting a window of time around 08:00 — i.e., *one second* — and identifying the single account ID closest to our temporal threshold when approaching from the left and the single account ID closest to our temporal threshold when approaching from the right based on the timestamp for when Octopus Energy sent each signed up customer their first (possibly only) opt-in notice.[39] We then summed these account IDs and divide by the value of two to construct our ID-based cutoff for treatment C_{ID}.

Our constructed ID-based threshold C_{ID} = "2,454,839". Account IDs ranged in size from "2" to "5,863,115" in our sample of 621,204 Octopus Energy customers who had signed up to participate in Saving Sessions by February 12 and for whom Octopus Energy tracked during the February 13, 2023 Saving Session. The constructed account ID of "2,454,839" was used as our "sharp" threshold for receipt of intraday notices. This threshold was "sharp" as only account IDs greater than or equal to "2,454,839" were regarded as receiving treatment (i.e., intraday notice).

It would have been impossible for Octopus Energy customers to manipulate their account ID as this would be tantamount to strategically modifying creation of their Octopus Energy account in relation to our threshold. Indeed, there was no way for Octopus Energy customers to influence their treatment assignment as C_{Time} and, by extension, C_{ID} were determined and only known by the authors of this research in relation to Octopus Energy's policy around the times at which customer communication is avoided (i.e., between 20:00 and 08:00). Furthermore, owing to our data on the time at which opt-in notices were sent by Octopus Energy, we know precisely which customers received treatment given our threshold — assuming, of course, that *sent* notices are actually *received and read* (i.e., full compliance).

Note that 11,673 of the 621,204 DFS-participating customers whom Octopus Energy tracked during the February 13, 2023 Saving Session were sent opt-in notices "overnight" (i.e., after 23:00 on February 12 but before 08:00 on February 13). We excluded these participants from our models owing to concerns about the stable unit treatment value assumption (i.e., "no hidden versions of treatments"; Gelman et al. (2020)). This exclusion was done under the assumption that overnight opt-in notices resulted in a fundamentally distinct treatment compared to the receipt of an intraday notice during working hours. And it resulted in a sample size of 609,531 DFS-participating customers.

[39] Notice "sent at" timestamps measured to the millisecond.

Removing Octopus Energy customers who received their first opt-in notice overnight necessarily resulted in what econometricians call a "donut-hole" regression discontinuity design (Barreca et al., 2011, 2016). In a donut-hole RDD, all observations with scores on the assignment variable A_i within some range immediately around the cutoff are excluded. Typically, excluding study units in this manner is done to address "heaping" (i.e., non-random clustering of observations at points along the observed range or "support" of the running variable). However, donut-hole RDD is an elegant means of handling overnight notices. That said, while notice timing *was* clustered due to Octopus Energy's batched dispersal of messages based on account IDs (see Section AI.7), we *did not* see evidence of heaping in relevant pre-treatment covariates across the range of our running variable using the visual diagnostic (Figure AF.16) recommended by Barreca et al. (2016).

Finally, because $C_{\text{Time}} = 08{:}00$, our sample was asymmetric around our ID-based cutoff C_{ID} to the left (Figure 8). Put alternatively, we exclude study units with values for the assignment variable A that fell immediately below C_{ID} (i.e., when A_i less than "2,454,839") whilst retaining observations to the immediate right of the cutoff.[40]

4.2 Regression Discontinuity Design: Methods

Exclusion of the 11,673 Octopus Energy customers sent overnight opt-in notices resulted in two groups divided by a sharp, account-ID-based discontinuity at C_{ID} = 2,454,839. Thus, $Pr(z_i = 1|A_i \geq C_{\text{ID}}) = 1$ *and* $Pr(z_i = 1|A_i < C_{\text{ID}}) = 0$. In this respect, treatment status z_i was fully determined by A_i.

Crucially, we assumed that Octopus Energy customers with values for the assignment variable A_i that fell on either side of the cut-off C_{ID} *within a restricted range or "bandwidth"* of account IDs h — i.e., $C_{\text{ID}} - h_{\text{Left}}$ and $C_{\text{ID}} + h_{\text{Right}}$ — were broadly "similar". That is to say, we assume that Octopus Energy customers very near to our cutoff had distributions of potential outcomes under treatment (i.e., y^1) and distributions of potential outcomes under no treatment (i.e., y^0) that were nearly equivalent conditional on A_i and possible confounders x (i.e., third variables determinant of both the observed outcome y and assignment A, and thus z).

This assumption is sometimes called "no confounders vary discontinuously across the threshold" Gelman et al. (2020, p. 438), however, it is a form of conditional ignorability Gelman et al. (2020, p. 438) in the following style:

$$y^0, y^1 \perp z \mid A,\ x \quad \forall \quad A \in (C_{\text{ID}} - h_{\text{Left}}, C_{\text{ID}} + h_{\text{Right}}). \tag{17}$$

There is generally a trade off between the plausibility of this assumption and the width of the bandwidth h, and thus the number of observations N used for model fitting (see Cattaneo et al. (Forthcoming) on justification of the "local randomization" RDD). Accordingly, and in keeping with standard practice across the academic literature, we relaxed the above assumption by narrowly focusing on estimating the causal effect of an intraday notice at the ID-based threshold — i.e., the causal effect when $A_i = C_{\text{ID}}$ — which is a kind of local average treatment effect (LATE).

Specifically, we estimated an ordinary-least-squares (OLS) regression model with the following general form:

$$\begin{gathered} y_i = \beta_0 + \beta_1 z_i + \beta_2 (A_i - C_{\text{ID}}) + \beta_3 (z_i \times (A_i - C_{\text{ID}})) + X_i \vec{\beta} + \epsilon_i \\ \forall \quad A_i \in (C_{\text{ID}} - h_{\text{Left}}, C_{\text{ID}} + h_{\text{Right}}), \end{gathered} \tag{18}$$

where C_{ID} = 2,454,839, z was our binary treatment indicator (i.e., intraday notice versus day-ahead notice) which equaled the value of one when $A_i \geq C_{\text{ID}}$, and $(A_i - C_{\text{ID}})$ was the assignment variable (i.e., account ID)

[40] There have been important methodological advances around regression discontinuity with multiple cut-offs (see Cattaneo et al. (Forthcoming)). However, we opted for a simpler analysis by limiting our attention to a comparison of intraday versus day-ahead opt-in notices. We did so as this comparison is most relevant to practical applications of our research by energy retailers and transmission system operators. This was because these institutions are likely to limit their interaction with consumers during unsociable, non-working hours.

relative to the cutoff. This relative quantity equals zero when $A_i = C_{\text{ID}}$ such that $z_i = 1$ when $(A_i - C_{\text{ID}}) \geq 0$. Furthermore, X is a $1 \times p$ matrix containing p pre-treatment covariates and/or confounders x believed to jointly determine both the potential outcomes (y^0, y^1) and A for study units $i \in N$ (Table AT.25) — with h_{Left} and h_{Right} defining the bandwidth of account IDs used for model fitting.

Moreover, (β_0) and $(\beta_0+\beta_1)$ are, respectively, the expected value of the response y at the threshold for customers for whom $z_i = 0$ and for whom $z_i = 1$. Thus, β_1 is the LATE at the threshold — i.e., $(\beta_0 + \beta_1) - (\beta_0)$ or the expected difference or "jump" in the outcome between customers for whom $z_i = 1$ and customers for whom $z_i = 0$.

In Section AI.10, we discuss the limited (i.e., restricted) range of our outcome variables in relation to our use of OLS regression. And, in Section AI.7.2, we also discuss our general model specification and the technique we use to estimate optimal, *asymmetric* bandwidths h_{Left} and h_{Right} given our use of donut-hole RDD. Optimal bandwidths are specific to each of our two outcome variables (i.e., session consumption and session participation). For session consumption, the lower bandwidth bound $(C - h_{\text{Left}})$ roughly corresponded to 22:57 on the 12th and the upper bandwidth bound $(C + h_{\text{Right}})$ roughly corresponded to 11:01 on the 13th (Figure 8 and Figure AF.16). For session participation, the lower and upper bounds roughly corresponded to 22:29 on the 12th and 10:52 on the 13th.

4.3 Regression Discontinuity Design: Results

Tables 6a and 6b present the causal effect (i.e., LATE) of receiving an intraday opt-in notice (i.e., $\hat{\beta}$ Intraday Opt-in Notice) from models of total consumption (kWh) during, and formal agreement to participate in, the 10th Saving Session on February 13.

Our simplest model of consumption used our standard pair of bandwidths (i.e., h_{Left} and h_{Right}) and excluded pre-treatment covariates. The OLS estimate of the LATE in this model indicated that Octopus Energy customers sent an intraday notice used, on average, 0.06 kWh more ($\hat{\beta}$ = 0.060; 95% CI = $[0.011, 0.108]$; p-value = 0.016) during the Saving Session compared to Octopus Energy customers sent a day-ahead notice — the latter of whom had an estimated average consumption of 0.566 kWh ($\hat{\beta}$ Intercept = 0.566; 95% CI = $[0.519, 0.613]$; p-value < 0.001). Put alternatively, there was compelling evidence to reject the null hypotheses of no association between our notice-based treatment and consumption, where our simplest models indicate that consumption on the part of customers sent a day-of notice was higher by 10.6% (i.e., 0.06 ÷ 0.566).

Note, however, that the OLS estimate of the LATE was sensitive to model specification. Specifically, it was attenuated in the presence of pre-treatment covariates ($\hat{\beta}$ = 0.042; 95% CI = $[0.002, 0.083]$; p-value = 0.041). In contrast, when we expanded our bandwidth by a factor of 1.5, the LATE increases in magnitude ($\hat{\beta}$ = 0.109; 95% CI = $[0.013, 0.204]$; p-value = 0.025). Nevertheless, when we expanded our bandwidth by a factor of two, we no longer observed compelling evidence to suggest that our notice-based treatment affected consumption ($\hat{\beta}$ = -0.021; 95% CI = $[-0.087, 0.046]$; p-value = 0.542).

Thus, there is compelling evidence to suggest that being sent an intraday notice had a positive causal impact on energy consumption during the 10th Saving Session, albeit only for a narrower set of customers receiving notice within a more restricted temporal window.

As for participation Table 6b, our linear probability models failed to provide sufficient evidence to reject the null hypotheses of no association between intraday notice and participation in the 10th Saving Session. This null result was found in our simplest model ($\hat{\beta}$ = 0.002; 95% CI = $[-0.013, 0.018]$; p-value = 0.767), our expanded model ($\hat{\beta}$ = -0.014; 95% CI = $[-0.029, 0.001]$; p-value = 0.067), and our models that expanded our bandwidth by a factor of 1.5 ($\hat{\beta}$ = -0.019; 95% CI = $[-0.039, 0.001]$; p-value = 0.068) and a factor of 2 ($\hat{\beta}$ = -0.025; 95% CI = $[-0.054, 0.004]$; p-value = 0.093).

Table 6: Results for Regression Discontinuity Design.

MSE-Optimal Bandwidth ($\mathbf{h}_{\text{Left}}$, $\mathbf{h}_{\text{Right}}$)	($\mathbf{h}_{\text{L}}$;, $\mathbf{h}_{\text{R}}$)	($\mathbf{h}_{\text{L}}$, $\mathbf{h}_{\text{R}}$)	($\mathbf{h}_{\text{L}\div 1.5}$, $\mathbf{h}_{\text{R}\times 1.5}$)	($\mathbf{h}_{\text{L}\div 2}$, $\mathbf{h}_{\text{R}\times 2}$)	$\mathbf{h}_{\text{L}}$ **Only**
$\hat{\beta}$ **Intercept**	0.566 (0.024)	0.588 (0.020)	0.532 (0.048)	0.653 (0.034)	0.613 (0.006)
$\hat{\beta}$ **Intraday Opt-in Notice**	0.060 (0.025)	0.042 (0.021)	0.109 (0.049)	-0.021 (0.034)	—
$\hat{\beta}$ **Intraday Opt-in Notice (07:59:59, 09:00:00]**	—	—	—	—	0.009 (0.008)
$\hat{\beta}$ **Intraday Opt-in Notice (09:00:00, 10:00:00]**	—	—	—	—	0.012 (0.008)
$\hat{\beta}$ **Intraday Opt-in Notice (10:00:00, 11:00:00]**	—	—	—	—	0.026 (0.007)
$\hat{\beta}$ **Intraday Opt-in Notice (11:00:00, 12:00:00]**	—	—	—	—	0.037 (0.007)
$\hat{\beta}$ **Intraday Opt-in Notice (12:00:00, 13:00:00]**	—	—	—	—	0.056 (0.008)
Pre-treatment Covariates?	No	Yes	Yes	Yes	Yes
Observations	78,724	69,168	96,477	123,280	350,361
Estimator	OLS	OLS	OLS	OLS	OLS
Heteroscedasticity-Consistent Std. Errors (HC0)	True	True	True	True	True
$R^2_{\text{Adj.}}$	0.000	0.309	0.305	0.985	0.961

(a) Models of total consumption (kWh) during the 10th Saving Session.

MSE-Optimal Bandwidth ($\mathbf{h}_{\text{Left}}$, $\mathbf{h}_{\text{Right}}$)	($\mathbf{h}_{\text{L}}$;, $\mathbf{h}_{\text{R}}$)	($\mathbf{h}_{\text{L}}$, $\mathbf{h}_{\text{R}}$)	($\mathbf{h}_{\text{L}\div 1.5}$, $\mathbf{h}_{\text{R}\times 1.5}$)	($\mathbf{h}_{\text{L}\div 2}$, $\mathbf{h}_{\text{R}\times 2}$)	$\mathbf{h}_{\text{L}}$ **Only**
$\hat{\beta}$ **Intercept**	0.563 (0.007)	0.566 (0.006)	0.566 (0.010)	0.572 (0.015)	0.576 (0.003)
$\hat{\beta}$ **Intraday Opt-in Notice**	0.002 (0.008)	-0.014 (0.008)	-0.019 (0.010)	-0.025 (0.015)	—
$\hat{\beta}$ **Intraday Opt-in Notice (07:59:59, 09:00:00]**	—	—	—	—	-0.026 (0.005)
$\hat{\beta}$ **Intraday Opt-in Notice (09:00:00, 10:00:00]**	—	—	—	—	-0.036 (0.005)
$\hat{\beta}$ **Intraday Opt-in Notice (10:00:00, 11:00:00]**	—	—	—	—	-0.045 (0.004)
$\hat{\beta}$ **Intraday Opt-in Notice (11:00:00, 12:00:00]**	—	—	—	—	-0.055 (0.004)
$\hat{\beta}$ **Intraday Opt-in Notice (12:00:00, 13:00:00]**	—	—	—	—	-0.053 (0.005)
Pre-treatment Covariates?	No	Yes	Yes	Yes	Yes
Observations	99,678	88,422	104,422	125,131	377,569
Estimator	OLS	OLS	OLS	OLS	OLS
Heteroscedasticity-Consistent Std. Errors (HC0)	True	True	True	True	True
$R^2_{\text{Adj.}}$	0.001	0.224	0.220	0.219	0.215

(b) Models of the probability of opting into the 10th Saving Session.

Note: The table presents parameter estimates and standard errors (parentheses) for the LATE ($\hat{\beta}$ Intraday Opt-in Notice [Ref Day-ahead Notice]), and the expected average outcome in the control group ($\hat{\beta}$ Intercept) from models fit to subsets of our Saving Session data using asymmetric bandwidths h_{Left} and h_{Right} optimized to reduce mean-squared error (MSE) (see Section AI.7.2). Hour-specific *ATEs* ($\hat{\beta}$ Intraday Opt-in Notice (Time Range]) are from a model fit to a subset of our data obtained using only the left MSE-optimal bandwidth. Results rounded to three decimal places. See Long and Ervin (2000) for a discussion and comparison heteroscedasticity-consistent covariance matrices. For complete results depicting all covariates, see Tables AT.26 and AT.27. See Table AT.25 for descriptive statistics and reference categories.

Contextualizing Demand Reduction. Before advancing, we also briefly considered treatment-effect heterogeneity in relation to the timing of intraday notice using a pair of ancillary models given in the right-most columns of Tables 6a and 6b. Recall that our binary indicator for intraday notice covered all notices sent roughly between 08:00 and 22:30 on February 13 (Figure 8). This leads to a comparison between customers sent notice within thirty minutes of 23:00 the day prior (i.e., the control group; Figure 8) and customers sent day-of notice over multiple hours across the morning of the 10th Saving Session. Accordingly, we explored whether our results might be consistent throughout the morning-to-afternoon period by fitting two additional models wherein we swapped our singular binary indicator for intraday notice with a series of binary indicators for the 60-min period within which day-of

Table 7: RDD-based LATE of intraday notice on total consumption (kWh) during the 10th Saving Session versus the ITT and the LATE for demand reduction during the same event obtained using DiDs.

RDD LATE Conditional on Pre-Treatment Covariates?	**No**	**Yes**
$\hat{\beta}_{\text{LATE}}$ **Intraday Opt-in Notice [Ref. Day-ahead Notice]**	0.060 (0.025)	0.042 (0.021)
$\hat{\beta}_{\text{ITT}}$ **Signed-up [Ref. Signed Up Never]**	-0.151 (0.003)	-0.151 (0.003)
$\hat{\beta}_{\text{LATE}}$ **Intraday Opt-in Notice as % of** $\hat{\beta}_{\text{ITT}}$ **Signed-up**	-39.74%	-27.81%
$\hat{\beta}_{\text{LATE (Sign-up)}}$ **Octopus Customers [Ref. Bulb Customers]**	-0.168 (0.005)	-0.168 (0.005)
$\hat{\beta}_{\text{LATE}}$ **Intraday Opt-in Notice as % of** $\hat{\beta}_{\text{LATE (Sign-up)}}$ **Octopus Customers**	-35.71%	-25.00%

Note: We multiplied the Session-specific DiD coefficients and their standard errors from Table AT.2 by two as the Saving Session on February 13, 2023 lasted two half-hours whereas the DiDs relate to average half-hourly electricity consumption. We then calculated the size of the effect of "intraday" notice we identified from our RDD (from our model without covariates, and with) as a percent of the demand reduction identified by the DiDs. The Signed Up Early versus Late DiD does not produce Session-specific estimates for Sessions in February and March 2022, as the control group for the Signed Up Early versus Late DiD comes from customers who joined in February and March 2022; for this reason, we only show the results as a percent of our Signed Up Early versus Never and Octopus versus Bulb DiDs. We found that depending on the model, the impact of intraday notice was 25-39% of the overall Sessions demand reduction signed-up customers achieved – dampening but not eliminating customers' flexibility response.

notice was sent (i.e., every hour from 08:00 and 13:00, the latter of which was the cut-off for the sending of opt-in notice on February 13).

In so doing, we necessarily approached our RDD through the lens of a standard observational study wherein we wished to estimate a causal effect by adjusting for all confounders (Gelman et al., 2020, p. 437). Thus, we assumed that, conditional on an array of pre-treatment covariates — most importantly customer tenure (Figure AF.16) — assignment to the control group or to the hour-specific treatment groups was independent of the potential outcomes within a restricted range of our running variable (See Equation (17) as well as Cattaneo et al. (Forthcoming) on "local randomization"). Put simply, we assumed ignorability of the hour-specific notice-based treatments conditional on tenure for a limited range of account IDs, amongst other factors.

We estimated our two ancillary models by filtering our data using only our lower bandwidth bound ($C_{\text{ID}} - h_{\text{Left}}$) such that all customers with account IDs A_i greater than ($C_{\text{ID}} - h_{\text{Left}}$) were used for model fitting. Thus, we continued to use as a control group customers sent notice between roughly 22:30 and 23:00 on February 12.

As this form of RDD focused on the average difference between the control and treatment groups (i.e., ATEs) as opposed to the causal effect at the cutoff C_{ID} (i.e., the LATE), it makes no functional form assumption (Cattaneo et al., Forthcoming), and standard analytic techniques can be used (Cattaneo et al., Forthcoming). Accordingly, the two ancillary models were estimated after dropping from the linear predictor account ID and the multiplicative interaction between intraday notice and account ID.

Keeping all of this in mind, the hour-specific ATEs (rightmost columns in Tables 6a and 6b) were generally consistent with the OLS estimates of the LATE from our main models in terms of their sign, where the hour-specific ATEs in both our model of consumption and our model of participation generally grow in magnitude over time.

In Tables 7 and 8, we contextualized our overall and our hour-specific treatment effects for consumption (Table 6a) using results from our DiD designs (Section 2.4.2). Specifically, we show the percent of in-Session demand reduction represented by these quantities which we derived by dividing our RDD's LATE for intraday notice, as well as the separate hour-specific ATEs, by the demand reduction associated with sign up that we estimated using our DiD designs(Table AT.2).[41] And, depending on the model, we found that the impact of intraday notice was 25-39% of the overall in-Sessions demand reduction signed-up customers achieved according to our DiD designs (Table 7) – dampening but far from eliminating customers' flexibility response. That said, we note that we prefer the RDD-derived LATE of intraday notice estimated while adjusting for possible confounds ($\hat{\beta}$ = 0.042; 95% CI = $[0.002, 0.083]$; p-value = 0.041) which we expect to add precision and improve the model's internal validity. And, when we divide the this point estimate by the LATE for signing up to taker part in DFS events obtained using our third DiD strategy (Octopus Customers vs. Bulb Customers), we see a 25% dampening of demand reduction.[42] As for the hour-by-hour ATEs as a percent of overall DiD-estimated demand reduction (Table 8), we found a notice period elasticity of 0.7 to 0.8, depending on which model we used to estimate in-Session demand reduction from signed up customers. In other words, for each 1% reduction in notice period, we saw a 0.7 to 0.8% dampening of the demand response customers achieve.

Table 8: Hour-specific elasticities (regression discontinuity design versus DiD designs).

Hours Before Session Start	$\hat{\beta}_{\text{ATE}}$ **Intraday Notice as % of** $\hat{\beta}_{\text{ITT}}$ **Signed-up**	**% Notice Period Reduced**	**Elasticity**
08:00 to 09:00 (9.5 hrs before)	-5.96% (0.009 [0.008] ÷ -0.151 [0.003])	0.00% ((9.5 - 9.5) ÷ 9.5)	—
09:00 to 10:00 (8.5 hrs before)	-7.95% (0.012 [0.008] ÷ -0.151 [0.003])	10.53% ((9.5 - 8.5) ÷ 9.5)	-0.75 (-7.95% ÷ 10.53%)
10:00 to 11:00 (7.5 hrs before)	-17.22% (0.026 [0.007] ÷ -0.151 [0.003])	21.05% ((9.5 - 7.5) ÷ 9.5)	-0.82 (-17.22% ÷ 21.05%)
11:00 to 12:00 (6.5 hrs before)	-24.50% (0.037 [0.007] ÷ -0.151 [0.003])	31.58% ((9.5 - 6.5) ÷ 9.5)	-0.78 (-24.50% ÷ 31.58%)
12:00 to 13:00 (5.5 hrs before)	-37.09% (0.056 [0.008] ÷ -0.151 [0.003])	42.11% ((9.5 - 5.5) ÷ 9.5)	-0.88 (-37.09% ÷ 42.11%)

(**a**) Elasticity given demand reduction estimated using the first DiD design (Signed Up Early vs. Signed Up Never).

Hours Before Session Start	$\hat{\beta}_{\text{ATE}}$ **Intraday Notice as % of** $\hat{\beta}_{\text{LATE}}$ **Octo. Cust.**	**% Notice Period Reduced**	**Elasticity**
08:00 to 09:00 (9.5 hrs before)	-5.37% (0.009 [0.008] ÷ -0.168 [0.005])	0.00% ((9.5 - 9.5) ÷ 9.5)	—
09:00 to 10:00 (8.5 hrs before)	-7.16% (0.012 [0.008] ÷ -0.168 [0.005])	10.53% ((9.5 - 8.5) ÷ 9.5)	-0.68 (-7.16% ÷ 10.53%)
10:00 to 11:00 (7.5 hrs before)	-15.51% (0.026 [0.007] ÷ -0.168 [0.005])	21.05% ((9.5 - 7.5) ÷ 9.5)	-0.74 (-15.51% ÷ 21.05%)
11:00 to 12:00 (6.5 hrs before)	-22.08% (0.037 [0.007] ÷ -0.168 [0.005])	31.58% ((9.5 - 6.5) ÷ 9.5)	-0.70 (-22.08% ÷ 31.58%)
12:00 to 13:00 (5.5 hrs before)	-33.41% (0.056 [0.008] ÷ -0.168 [0.005])	42.11% ((9.5 - 5.5) ÷ 9.5)	-0.79 (-33.41% ÷ 42.11%)

(**b**) Elasticity given demand reduction estimated using the third DiD design (Octopus Customers vs. Bulb Customers).

Note: Here we report the hour-by-hour notice period "elasticity" using the hour-specific treatment effects $\hat{\beta}_{\text{ATE}}$ Intraday Opt-in Notice (Time Range] (standard errors in brackets) from Table 6, the Session-specific intent-to-treat effect $\hat{\beta}_{\text{ITT}}$ Signed Up Early (Ref. Signed Up Never) using our first DiD strategy (Table AT.2) and the Session-specific LATE $\hat{\beta}_{\text{LATE (Sign-up)}}$ Octopus Customers (Ref. Bulb Customers) from our third DiD strategy (Table AT.2). To calculate the "elasticity" for a given sixty-minute period, we first divide the corresponding hour-specific ATE by the demand reduction reflected in either the ITT effect (Table 8a; i.e., the causal effect of eligibility for Sessions participation on average half-hourly consumption during the February 13 Saving Session) or the LATE (Table 8b; i.e., the causal effect of signing-up to the DFS itself on average half-hourly consumption during the February 13 Saving Session among customers who's sign-up behaviour was influenced by the DFS invitation). After multiplying by 100, this division (second column) provides us with the percent reduction in demand during the February 13 Saving Session associated with receiving notice one additional hour after the 08:00 to 09:00 period. Next (third column), using 9:00 as a baseline, we calculate the percentage reduction in the overall notice period (08:00 to 13:00) incurred by receiving notice one additional hour after the 08:00 to 09:00 period. Then, to calculate the hour-specific "elasticity" (fourth column), we simply divide the percentage of the estimated reduction in demand that is associated with receiving notice two, three, four, of five hours after the 08:00 to 09:00 period (according to our RDD) by the percentage reduction in the overall notice period due to the elapsing of two, three, four, of five hours. Session-specific DiD effects and their standard errors (brackets; second column) are multiplied by two as the Saving Session on February 13, 2023 lasted two half-hours.

[41] Recall that our second DiD design (i.e., Signed Up Early vs Signed Up Late) does not produce Session-specific estimates for Saving Sessions in February and March 2023. For this reason, we only show the ATE as a percent of the causal effects obtained using our Signed Up Early vs. Never DiD design and our Octopus vs. Bulb DiD design.

[42] There is extremely minimal risk of Bulb customers choosing Bulb as their supplier in anticipation of DFS. In addition recall from Section 2.4.2 that, for multiple reasons, Bulb customers are a "natural" counterfactual group for Octopus Energy customers invited to participate in the DFS.

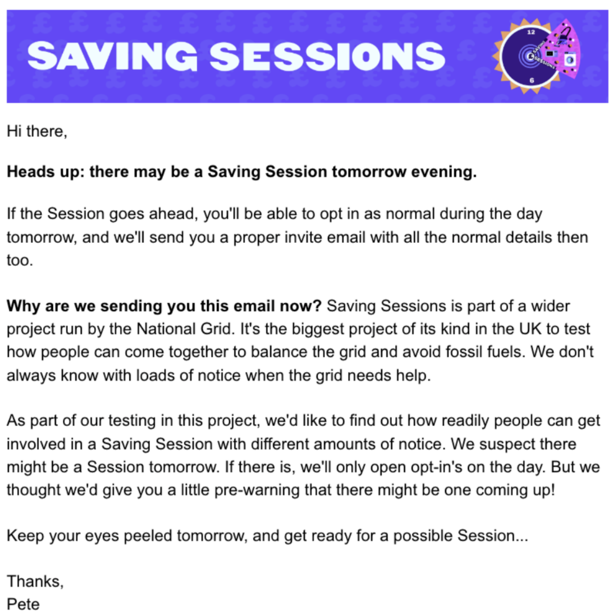
4.4 Timing and Type of Supplementary Saving Session Notices

Here we further consider the causal impact of opt-in notice timing using a randomized field experiment based on an inversion of the customer-messaging set-up used for the February 13 quasi-experiment discussed in Section 4.1.

Our field experiment concerned 650,809 Octopus Energy customers who had signed up to participate in Saving Sessions by March 14 and whom Octopus Energy tracked during the Saving Session on March 15 (18:30 to 19:30). Unlike the other Saving Sessions throughout the Winter of 2022-23, all 650,809 customers were sent *intraday* opt-in notices for the March 15 Session. Accordingly, for our field trial, we randomly selected a subset of customers to receive *supplementary messaging* and, in some cases, a *supplementary price incentive* on top of the standard intraday notice.

Of the 650,809 Octopus Energy customers, 19,182 were randomly assigned to receive a *day-ahead* "heads-up" email on March 14 (Figure 9). We refer to this notice as a "heads-up" email as those in receipt of this message *could not* use it to formally agree to participate in the Saving Session on March 15 (c.f. opt-in notices). Indeed, the heads-up email only informed DFS-participating Octopus customers about the upcoming Saving Session and its general importance. Distribution of the day-ahead heads-up email was managed using a third-party platform (i.e., "SendGrid") that is operationally distinct from Octopus Energy's internal customer-messaging platform used to send the intraday notices (Section AI.7).

Figure 9: Day-ahead "heads-up" email sent to customers to raise awareness about an upcoming Saving Session.



SAVING SESSIONS

Hi there,

Heads up: there may be a Saving Session tomorrow evening.

If the Session goes ahead, you'll be able to opt in as normal during the day tomorrow, and we'll send you a proper invite email with all the normal details then too.

Why are we sending you this email now? Saving Sessions is part of a wider project run by the National Grid. It's the biggest project of its kind in the UK to test how people can come together to balance the grid and avoid fossil fuels. We don't always know with loads of notice when the grid needs help.

As part of our testing in this project, we'd like to find out how readily people can get involved in a Saving Session with different amounts of notice. We suspect there might be a Session tomorrow. If there is, we'll only open opt-in's on the day. But we thought we'd give you a little pre-warning that there might be one coming up!

Keep your eyes peeled tomorrow, and get ready for a possible Session...

Thanks,
Pete

Note: Octopus Energy sent day-ahead notices to a randomly chosen group of customers sent on March 14, 2023. This email effectively let them know of the possibility of an incoming Saving Session on the next day; they were also required to opt-in to the session held on March 15, 2023.

Furthermore, 19,220 of the 650,809 Octopus Energy customers were randomly assigned to receive an *intraday* "reminder" SMS text message on March 15. And these 19,220 customers were made eligible for a performance-related bonus price incentive of 1,000 "OctoPoints" worth £1.25. Like the heads-up email, the SMS reminder text raised awareness of the upcoming Saving Session. However, customers in receipt of a text may have already received primary notice (hence, "reminder"). Furthermore, the SMS text disclosed the level of bonus on offer subject to positive session performance.[43] Owing to its short length, the SMS text did not reference the general importance

[43]Performance was deemed positive when a customer's in-Session consumption was less than their P376 ("clipped") baseline consumption with in-day adjustment.

of Saving Sessions. The exact SMS text was as follows:

> "SAVING SESSION TODAY 1830-1930. SPOT PRIZE: Octobot has chosen you at random to win 1000 OctoPoints if you save energy in tonight's Session. Opt in before 1830!"

Thus, our field experiment has three experimental conditions:

1. Control Group ($N = 627{,}155$): Intraday Opt-in Notice Only
2. Treatment Group 1 ($N = 19{,}182$): Intraday Opt-in Notice *plus* Day-ahead "Heads-up" Email
3. Treatment Group 2 ($N = 19{,}220$): Intraday Opt-in Notice *plus* Intraday "Reminder" SMS Text *plus* Eligibility for £1.25 Bonus

Two factors made random assignment of our second treatment imperfect. First, some customers assigned to the SMS-plus-bonus condition had disallowed SMS communications from Octopus Energy. Second, we were unable to send SMS texts to every customer assigned to the SMS-plus-bonus condition who *had* allowed SMS messages from Octopus Energy.

Specifically, we were limited to sending intraday SMS texts to a *maximum* of 5,000 Octopus Energy customers. In total, 4,731 of the 19,220 Octopus Energy customers assigned to the SMS-plus-bonus condition *did not* allow SMS texts from Octopus Energy. Of the remaining 14,489 customers who did allow SMS communications, 4,472 were randomly sub-sampled to receive an intraday SMS reminder. Thus, there were 14,748 customers (i.e., 4,731 + 10,017) assigned the second treatment who did not actually receive the second treatment. Nevertheless, all 19,220 customers assigned to the SMS-plus-bonus condition were made eligible to receive the bonus price incentive regardless of whether they disallowed SMS communications from Octopus Energy and irrespective of whether they were a part of the random sub-sample. In these respects, our second treatment suffers from imperfect compliance.

We were interested in the causal effect of actually receiving an intraday SMS text — again assuming, similarly to our RDD, that *sent* notices are actually *received*. Thus, our binary indicator for the second treatment only reflected the 4,472 Octopus Energy customers who had allowed SMS communication from Octopus Energy and who were sent an intraday SMS reminder after random sub-sampling. Consequently, the 4,731 customers who had disallowed SMS communication and the 10,017 customers who had allowed SMS communication but who were not randomly sub-sampled were only included in our binary indicator for *eligibility* for our second treatment.

Note well that these 14,748 customers only received the intraday email. And they were not made aware of their eligibility for the bonus price incentive unless they met the bonus criterion by the end of the Saving Session — where winnings were disclosed after the Session. If these individuals *had* been informed, it would represent a distinct form of treatment and it would have been prudent to create a fourth treatment group for "intraday notice plus bonus eligible". But this was not the case.

Keeping all of this in mind, we drew the following contrasts across the 650,809 Octopus Energy customers who had signed up to participate in DFS events by March 14 and for whom Octopus Energy tracked during the March 15 Saving Session as a part of our field experiment:

1. Intraday Opt-in Notice + Day-ahead "Heads-up" Email (Treatment Group 1) vs. Intraday Opt-in Notice Only (Control Group)

2. Intraday Opt-in Notice + Intraday "Reminder" SMS Text [Actually Received] + Eligibility for £1.25 Bonus (Treatment Group 2) vs. Intraday Opt-in Notice Only (Control Group)

Given non-compliance, we adopted the framing of a randomized encouragement design (RED) — i.e., a type of experimental setup wherein variation in some difficult-to-directly-manipulate treatment is induced using a source of random variation (i.e., the random "encouragement") that is related to the difficult-to-directly-manipulate treatment *and not* related to the outcome of interest. Here, our random "encouragement" is our *original*, explicit random assignment to the SMS-plus-bonus condition.

Note, due to non-compliance, we could only estimate the complier average causal effect (CACE) for the SMS-plus-bonus condition. The CACE — itself a kind of LATE — is the causal estimand for customers whose receipt of an intraday SMS and awareness of their eligibility for the bonus price incentive could be altered by our randomization. This is distinct from the average treatment effect (ATE) which we estimated for the day-ahead-email condition alongside the CACE using instrumental variable (IV) estimation. We further explain and justify our RED framing and our use of IV estimation for our field experiment in Section AI.8.2.

4.5 Field Trial: Methods

We estimated the ATE and the CACE for our supplementary-notice conditions using a two-stage ordinary least-squares (2SLS) procedure (Gelman et al., 2020, Greene, 2019, Wooldridge, 2010). The first stage (i.e., the first model) was an OLS regression of our treatment with imperfect compliance T_{SMS} (i.e., a binary indicator for the 4,472 customers who received the SMS-plus-bonus treatment) conditional on our binary random encouragement Z_{SMS} (i.e., an indicator for the 19,220 customers merely assigned to the SMS-plus-bonus condition). The second stage (i.e., the second model) was an OLS regression of our outcome y (i.e., in-Session consumption or Session participation) conditional on the predicted value of T_{SMS} from our first model, i.e., $\hat{T}_{\text{SMS}}$.

Formally, this combined model for the ATE for the day-ahead-email condition $T_{\text{Day-ahead Email}}$ and the *CACE* for the SMS-plus-bonus condition T_{SMS}, which we estimate using a joint procedure to correct the standard errors in the second stage (Gelman et al., 2020), is as follows:

$$y_i = \beta_0 + \beta_1 T_{\text{Day-ahead Email},i} + \beta_2 \hat{T}_{\text{SMS},i} + X_i \vec{\beta} + \epsilon_{i,y} \tag{19a}$$

$$T_{\text{SMS},i} = \gamma_0 + \gamma_1 T_{\text{Day-ahead Email},i} + \gamma_2 Z_{\text{SMS},i} + X_i \vec{\gamma} + \epsilon_{i,T_{\text{SMS}}}, \tag{19b}$$

where the linear predictor for T_{SMS} (i.e., Equation (19b)) and the linear predictor for y_i (i.e., Equation (19a)) are the first- and second-stage equations, $T_{\text{Day-ahead Email},i}$ is the exogenous binary indicator for treatment one (i.e., day-ahead "heads-up" email), $T_{\text{SMS},i}$ is the endogenous binary indicator for treatment two *for Octopus Energy customers who received it*, $\hat{T}_{\text{SMS}}$ is its predicted value, and $Z_{\text{SMS},i}$ is our random instrument/encouragement (i.e., random assignment to the SMS-plus-bonus condition).

Moreover, X is a $1 \times p$ matrix containing p pre-treatment covariates and/or confounders x believed to jointly determine $Z_{\text{SMS},i}$ and the *response* potential outcomes (y^0, y^1) or $Z_{\text{SMS},i}$ and the *treatment* potential outcomes $(T^0_{\text{SMS}}, T^1_{\text{SMS}})$ for study units $i \in N$. Accordingly, $\vec{\beta}$ and $\vec{\gamma}$ are p-length vectors of coefficients relating the pre-treatment covariates/confounders to y and T_{SMS}, respectively. Note that we include $T_{\text{Day-ahead Email},i}$ in Equation (19b) in keeping with typical applications of 2SLS whereby the first- and second-stage linear predictors are identical save for instrument and the associated treatment-with-non-compliance.

Given this set-up, β_1 in Equation (19a) is the ATE of the day-ahead condition, β_2 in Equation (19a) is the CACE of the SMS-plus-bonus condition, and γ_2 in Equation (19b) captures the expected compliance rate.

Alongside our interrelated regression equations (Equation (19)), we consider a single-equation model akin to Equation (19a) wherein we replaced the predicted value for actually receiving the SMS-plus-bonus treatment with a binary indicator for randomized assignment to the SMS-plus-bonus condition. This is the "reduced form" version of our two-equation model (see Wooldridge (2010, p. 90-91) and Angrist (2006, p. 32-33)) that provides us with the intent-to-treat (ITT) effect (Gelman et al., 2020, p. 426) — i.e., the effect of merely being eligible for SMS-based treatment based on our random assignment.

Table 9: Models of total consumption (kWh) during the 12th Saving Session (field trial).

Equation Response Variable	Reduced Fm. Consumption	Reduced Fm. Consumption	Reduced Fm. Consumption	1$^{\text{st}}$ Stage SMS + £	2$^{\text{nd}}$ Stage Consumption	2$^{\text{nd}}$ Stage Consumption	2$^{\text{nd}}$ Stage Consumption
$\hat{\beta}$ **Intercept**	0.655 (0.001)	0.656 (0.001)	0.662 (0.001)	-0.000 (0.000)	0.655 (0.001)	0.656 (0.001)	0.662 (0.001)
$\hat{\beta}$ **Day-ahead Email**	-0.021 (0.006)	-0.018 (0.006)	-0.011 (0.005)	0.000 (0.000)	-0.021 (0.006)	-0.018 (0.006)	-0.011 (0.005)
$\hat{\beta}$ **Intraday SMS + £1.25 Ass.**	-0.007 (0.006)	-0.007 (0.006)	-0.007 (0.005)	0.228 (0.003)	—	—	—
$\hat{\beta}$ **Intraday SMS + £1.25**	—	—	—	—	-0.030 (0.025)	-0.033 (0.026)	-0.029 (0.021)
Pre-treatment Covariates?	No	No	Yes	Yes	No	No	Yes
Listwise Deletion?	No	Yes	Yes	Yes	No	Yes	Yes
Observations	638,242	540,395	540,395	540,395	638,242	540,395	540,395
Estimator	OLS	OLS	OLS	OLS	IV-2SLS	IV-2SLS	IV-2SLS
Hetero.-Consist. SEs (HC0)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$R^2_{\text{Adj.}}$	0.000	0.000	0.367	0.223	0.000	0.000	0.367
Partial F-Statistic				4728.325			
Exogeneity Test p-value					0.433	0.554	0.485

Note: The table presents parameter estimates and standard errors (parentheses) for the ATE ($\hat{\beta}$ Intraday Notice + Day-ahead Email [Ref Intraday Only]), the CACE ($\hat{\beta}$ Intraday Notice + Intraday SMS + £1.25 Incentive [Ref Intraday Only]), the ITT ($\hat{\beta}$ Intraday SMS + £1.25 Incentive Assigned), and the expected average outcome in the control group ($\hat{\beta}$ Intercept) from reduced form ordinary least-squares (OLS) regression models as well as the 1^{st} and 2^{nd} stages of two-stage-least-squares (2SLS) regression models. "Intraday SMS + £1.25 Incentive Assigned" is the singular instrumental variable (IV). Results rounded to three decimal places. See Long and Ervin (2000) for a discussion and comparison heteroscedasticity-consistent covariance matrices. H_0 for Wooldridge's regression text of exogeneity is that the endogenous variable "Intraday SMS + £1.25 Incentive Assigned" is exogenous. For complete results depicting all covariates, see Table AT.28. See Table AT.25 for descriptive statistics and reference categories.

Formally, our single-equation model for the ITT is defined as follows:

$$y_i = \theta_0 + \theta_1 T_{\text{Day-ahead Email},i} + \theta_2 Z_{\text{SMS Randomly Assigned},i} + X_i \vec{\theta} + \epsilon_i, \tag{20}$$

where θ_1 is the ATE for the day-ahead condition as in Equation (19a), θ_2 is the ITT which can be "scaled" (i.e., divided) by the compliance rate γ_2 in Equation (19b) to produce a Wald-like estimate of the CACE (i.e., β_2 in Equation (19a) (Gelman et al., 2020, p. 426), and X is a $N \times p$ matrix containing p pre-treatment covariates.

4.6 Field Trial: Results

Tables 9 and 11 present three causal effects from models of total consumption (kWh) during, and formal agreement to participate in, the 12th Saving Session on March 15. Specifically, we present OLS-based estimates of: (a) the ATE of receiving a day-ahead heads-up email (i.e., $\hat{\beta}$ Intraday Notice + Day-ahead Email); (b) the CACE of receiving an intraday SMS reminder plus being made bonus-eligible (i.e., $\hat{\beta}$ Intraday Notice + Intraday SMS + £1.25 Incentive); and (c) the intent-to-treat effect (ITT) for the SMS-plus-bonus condition (i.e., $\hat{\beta}$ Intraday Notice + Intraday SMS + £1.25 Incentive *Assigned*).

Our simplest 2^{nd} stage model of consumption excluded pre-treatment covariates (Table 9). And the OLS estimate of the ATE from this model indicates that Octopus Energy customers sent a supplementary day-ahead email used, on average, 0.021 kWh *less* ($\hat{\beta}$ = -0.021; 95% CI = $[-0.032, -0.010]$; p-value < 0.001) during the March 15 Saving Session compared to customers only sent an intraday notice. On average, among customers who only received day-of notice, consumption was estimated to be 0.655 kWh ($\hat{\beta}$Intercept $=$ 0.655; 95% CI $=$ $[0.653, 0.657]$; p-value

< 0.001). Accordingly, the ATE represents a 3.2% (i.e., $0.021 \div 0.655$) reduction in consumption over baseline. Thus, we obtained compelling evidence to reject the null hypotheses that the causal association between being sent a supplementary day-ahead email and in-Session consumption is equal to zero. Nevertheless, the OLS estimate of the ATE of heads-up notice was attenuated in the presence of pre-treatment covariates ($\hat{\beta}$ = -0.011; 95% CI $= [-0.021, -0.002]$; p-value $= 0.021$), where this attenuation does not appear to be wholly related to our dropping of customers with incomplete data for pre-treatment covariates (i.e., listwise deletion).

In Table 10, we contextualized our ATE of the day-ahead email on in-Session consumption (Table 9) using results from our DiD designs (Section 2.4.2). Specifically, we show the percent of in-Session demand reduction represented by this quantity which we derived by dividing our field trials's ATE for the day-ahead email by the demand reduction associated with sign up that we estimated using our DiD designs(Table AT.2) [44]. And, depending on model, we found that the impact of the day-ahead "heads-up" was 7-19% of the overall in-Sessions demand reduction signed-up customers achieved. As in Section 4.3, our preferred estimate of the ATE is the version from our models that adjusts for pre-treatment covariates.

In contrast to our email-based treatment, we found no compelling evidence to suggest that being sent an intraday SMS reminder while being made eligible for the bonus price incentive had a causal impact on consumption during the March 15 Saving Session. Specifically, the 1^{st} stage model captured the unsurprising association between our instrument and our SMS-based treatment with non-compliance ($\hat{\beta}$Intraday SMS + £1.25 Incentive Assigned = 0.228; 95% CI $= [0.221, 0.234]$; p-value < 0.001) which represents the expected proportion of compliers, conditional on pre-treatment covariates. Nevertheless, our 2^{nd} stage models used to estimate the CACE failed to provide sufficient evidence to reject the null hypotheses of no association between our SMS-based treatment and consumption during the 12th Saving Session. This null result is found in our simple 2^{nd} stage model ($\hat{\beta}$ Intraday SMS + £1.25 Incentive = -0.030; 95% CI $= [-0.078, 0.019]$; p-value $= 0.229$) and the variant with pre-treatment covariates ($\hat{\beta}$ = -0.029; 95% CI $= [-0.070, 0.013]$; p-value $= 0.172$). Along this line, we also failed to find compelling evidence to reject the null hypotheses that the OLS estimate of the ITT effect is zero — i.e., that mere eligibility for the SMS-plus-bonus treatment impacts in-Session consumption, conditional on pre-treatment covariates (i.e., $\hat{\beta}$ Intraday SMS + £1.25 Incentive Assigned = -0.007; 95% CI $= [-0.016, 0.003]$; p-value $= 0.172$). The OLS estimate of the ITT is unchanged when excluding pre-treatment covariates.

As for participation (Table 11), our simple and expanded 2^{nd} stage models (linear probability sub-models) both provided compelling evidence of a positive association between our treatments and the likelihood of opting into the March 15 Saving Session. Specifically, the OLS estimate of the ATE for the day-ahead-heads-up-email condition from our simplest model without listwise deletion ($\hat{\beta}$ = 0.026; 95% CI $= [0.019, 0.033]$; p-value < 0.001) indicated a roughly 6% (i.e., $0.026 \div 0.423$) increase in the the probability of opt-in over the control group, the latter of whom are estimated as opting-in with a probability of 0.443, on average (95% CI $= [0.442, 0.444]$; p-value < 0.001). Thus, there is compelling evidence to reject the null hypotheses that the causal association between being sent a supplementary day-ahead email and the probability of Session participation is equal to zero. This evidence persisted when we adjusted for pre-treatment covariates ($\hat{\beta}$ = 0.025 (2^{nd} Stage Model); 95% CI $= [0.018, 0.032]$; p-value < 0.001).

Further still, the OLS estimate for the CACE of the SMS-plus-bonus condition is, respectively, 0.102 (95% CI $= [0.019, 0.033]$; p-value < 0.001) and 0.098 (95% CI $= [0.064, 0.132]$; p-value < 0.001) in our simple and expanded models of the probability of participating in the 12th Saving Session. And, in both cases, there is compelling evidence to reject the null hypotheses that the causal association between participation and actually receiving SMS-based notice, while being made eligible for the cash bonus, was equal to zero. These results indicate that being sent a supplementary intraday SMS-based reminder while being made eligible for an additional price incentive caused

[44] Once again recall that our second DiD design (i.e., Signed Up Early vs Signed Up Late) does not produce Session-specific estimates for Saving Sessions in February and March 2023. For this reason, we only show the ATE as a percent of the causal effects obtained using our Signed Up Early vs. Never DiD design and our Octopus vs. Bulb DiD design.

Table 10: Trial-based ATE of the day-ahead "heads-up" email on total consumption (kWh) during the 12th Saving Session versus the ITT and the LATE for demand reduction during the same event obtained using DiDs.

Field Trial ATE Conditional on Pre-Treatment Covariates?	**No**	**Yes**
$\hat{\beta}_{\textbf{ATE}}$ **Intraday Notice + Day-ahead Email [Ref Intraday Notice Only]**	-0.021 (0.006)	-0.011 (0.005)
$\hat{\beta}_{\textbf{ITT}}$ **Signed-up [Ref. Signed Up Never]**	-0.111 (0.002)	-0.111 (0.002)
$\hat{\beta}_{\textbf{ATE}}$ **Intraday Notice + Day-ahead Email as % of** $\hat{\beta}_{\textbf{ITT}}$ **Signed-up**	18.95%	9.93%
$\hat{\beta}_{\textbf{LATE (Sign-up)}}$ **Octopus Customers [Ref. Bulb Customers]**	-0.166 (0.005)	-0.166 (0.005)
$\hat{\beta}_{\textbf{ATE}}$ **Intraday Notice + Day-ahead Email as % of** $\hat{\beta}_{\textbf{LATE (Sign-up)}}$ **Octopus Customers**	12.65%	6.63%

Note: We multiplied the DiD coefficients and SEs from Table AT.2 by two, as the Saving Session on March 15, 2023 lasted two half-hours (one hour), whereas the DiDs estimated the impact on half-hourly electricity consumption. We then calculated the size of the effect of the day-ahead "heads-up" we identified from our field experiment (from our model without covariates, and with) as a percent of the demand reduction identified by the DiDs. The Signed Up Early versus Late DiD did not produce Session-specific estimates for Sessions in February and March 2022, as the control group for the Signed Up Early versus Late DiD comes from customers who joined in February and March 2022; for this reason, we only show the results as a percent of our Signed Up Early versus Never and Octopus versus Bulb DiDs.Depending on the model, the impact of the day-ahead "heads-up" was 7-19% of the overall Sessions demand reduction signed-up customers achieved.

an increase in the probability of opt-in by 23% (i.e., 0.098 ÷ 0.422) over the probability of event participation in the control group ($\hat{\beta}$ Intercept = 0.422; 95% CI = $[0.421, 0.424]$; p-value < 0.001), conditional on pre-treatment covariates.[45] Along this line, we also found compelling evidence to reject the null hypotheses that the OLS estimate of the ITT effect is zero — i.e., that, conditional on covariates, mere eligibility for the SMS-plus-bonus treatment impacted the probability of Session participation (i.e., $\hat{\beta}$ Intraday SMS + £1.25 Incentive Assigned (Reduced Form Model) = 0.022; 95% CI = $[0.015, 0.029]$; p-value < 0.001). The ITT effect represents a 5% (i.e., 0.022 ÷ 0.422) increase in the probability of opt-in due to treatment eligibility.

[45] In summary, then, the SMS with incentive caused more customers to opt in to the event, but we did not see an effect of the SMS with incentive on electricity demand. This may be due to low power to measure the effect of the SMS with incentive on electricity demand. However, it may also be related to the incentive's design. The incentive gave a bonus to customers for turning down *at all*. This may have induced extra opt-in from marginal signed up customers in case the household had lower consumption than their baseline, perhaps even by chance. But it has no bearing on the incentives of customers already planning to reduce demand. In other words, it encourages extra participation on the extensive margin but not the intensive margin.

Table 11: Models of the probability of opting into the 12th Saving Session (field trial).

Equation Response Variable	Reduced Fm. Participation	Reduced Fm. Participation	Reduced Fm. Participation	1^{st} Stage SMS + £	2^{nd} Stage Participation	2^{nd} Stage Participation	2^{nd} Stage Participation
$\hat{\beta}$ **Intercept**	0.443 (0.001)	0.434 (0.001)	0.422 (0.001)	-0.000 (0.000)	0.443 (0.001)	0.434 (0.001)	0.422 (0.001)
$\hat{\beta}$ **Day-ahead Email**	0.026 (0.004)	0.026 (0.004)	0.025 (0.004)	0.000 (0.000)	0.026 (0.004)	0.026 (0.004)	0.025 (0.004)
$\hat{\beta}$ **Intraday SMS + £1.25 Ass.**	0.024 (0.004)	0.022 (0.004)	0.022 (0.004)	0.227 (0.003)	—	—	—
$\hat{\beta}$ **Intraday SMS + £1.25**	—	—	—	—	0.103 (0.016)	0.095 (0.017)	0.098 (0.016)
Pre-treatment Covariates?	No	No	Yes	Yes	No	No	Yes
Listwise Deletion?	No	Yes	Yes	Yes	No	Yes	Yes
Observations	650,809	551,494	551,494	551,494	650,809	551,494	551,494
Estimator	OLS	OLS	OLS	OLS	IV-2SLS	IV-2SLS	IV-2SLS
Hetero.-Consist. SEs (HC0)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$R^2_{\text{Adj.}}$	0.000	0.000	0.174	0.223	0.001	0.001	0.175
Partial F-Statistic				4814.807			
Exogeneity Test p-value					0.021	0.021	0.010

Note: The table presents parameter estimates and standard errors (parentheses) for the ATE ($\hat{\beta}$ Intraday Notice + Day-ahead Email [Ref Intraday Only]), the CACE ($\hat{\beta}$ Intraday Notice + Intraday SMS + £1.25 Incentive [Ref Intraday Only]), the ITT ($\hat{\beta}$ Intraday SMS + £1.25 Incentive Assigned), and the expected average outcome in the control group ($\hat{\beta}$ Intercept) from reduced form ordinary least-squares (OLS) regression models as well as the 1^{st} and 2^{nd} stages of two-stage-least-squares (2SLS) regression models. "Intraday SMS + £1.25 Incentive Assigned" is the singular instrumental variable (IV). Results rounded to three decimal places. See Long and Ervin (2000) for a discussion and comparison heteroscedasticity-consistent covariance matrices. H_0 for Wooldridge's regression text of exogeneity is that the endogenous variable "Intraday SMS + £1.25 Incentive Assigned" is exogenous. For complete results depicting all covariates, see Table AT.29. See Table AT.25 for descriptive statistics and reference categories.

5 Cost effectiveness and welfare impacts of DFS

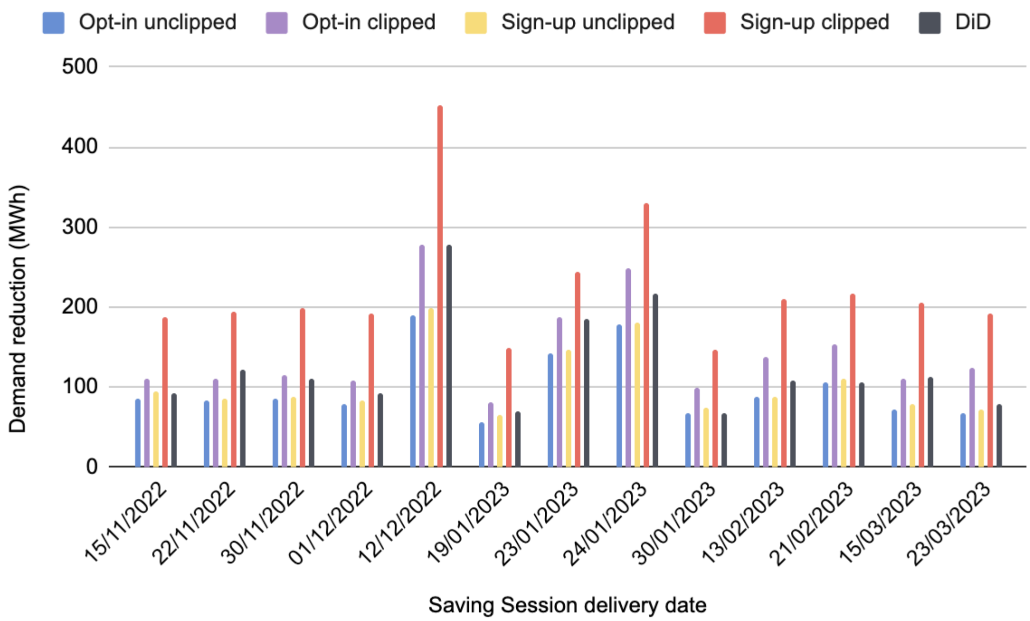
5.1 Total demand reduction from Saving Sessions

We estimated the total demand reduction caused by Saving Sessions using the LATE on sign-up from our Octopus versus Bulb DiD. We judged this estimate of the impact of sign-up on demand to be especially strong for reasons discussed in Section 2.4.2, namely that there is extremely minimal risk of Bulb customers choosing Bulb as their supplier in anticipation of DFS, making their presence in our dataset a useful "natural" counterfactual group for Octopus customers invited to participate in DFS. We calculated a separate effect for each Saving Session, interpretable as the impact of signing up to Saving Sessions (by the Session in question) on consumption per half-hour during the Session. We then multiplied this coefficient by the number of half-hours in the Session (usually two, but three for the Session on January 24, 2023; and four for the Session on December 12, 2022) and by the number of customers who had signed up Saving Sessions by that Session.[46]

We compare the demand reduction derived from our DiD to the demand reduction calculated through application of the P376 methodology prescribed by NGESO, described in Section 2.4.1. As noted in Section 3.2, clipping inflates demand reduction, but mostly when including all signed up customers in the calculation of demand reduction. Insofar as we consider our DiD results to be the relevant barometer of accuracy, the official methodology – clipped demand reduction among opted in customers only – overestimates demand reduction by ≈13%.

[46]Note that the validity of this multiplication relies on a simplifying assumption: that the customers in our sample had a demand reduction from sign-up that is representative of customers out of our sample (i.e., the customers excluded from analysis for having insufficient consumption data, moving home or supplier during the Saving Sessions period, or being a non-domestic customer).

Figure 10: Total demand reduction (MWh) according to various methodologies.



Note: We show demand reduction from Saving Session according to five potential methods: unclipped reduction among opt-ins only; clipped reduction among opt-ins only; unclipped reduction among all signed up customers; clipped reduction among all signed up customers; and reduction as measured by the LATE on sign-up in our Octopus versus Bulb DiD. The official methodology is the second: clipped reduction among opt-ins only. We found it is on average 13% higher than the result implied by our DiD. Meanwhile, interestingly, the unclipped reduction among all signed up customers was very close to our DiD result.

It was not clear to us why *unclipped* demand reduction among opted in customers is consistently lower than our DiD estimates. However, the unclipped version of the NGESO P376 methodology is a pre-post, and it was arguably just as likely to be downward as upward biased. In the following Sections 5.2 and 5.3, we use our Octopus versus Bulb DiD Session-specific estimates (as shown in Figure 10) as our preferred estimate of total demand reduction.

5.2 Cost effectiveness

In Great Britain, NGESO and distribution network operators procure services to ensure real-time balancing of generation and demand as well as ancillary services dealing with inertia management (a property of grid management related to resisting changes in system frequency), local voltage constraints, and constraints between transmission boundaries (LCP Delta, 2023). The total costs of these services in the winter of 2022/23 was £1,235M, 20% less than the previous winter, but still higher than previous years, according to LCPDelta's Balancing Costs Review, conducted on behalf of National Grid NGESO (LCP Delta, 2023).

These costs have increased in the past two years due to tighter system margins, high wind outputs which cause higher transmission constraint management costs between regions with high wind and other parts of Great Britain, and high gas prices. Balancing Mechanism costs were higher in 2021/22 and 2022/23 than ever before due to both greater volume of balancing actions required and higher prices per balancing action. For example, the volume weighted accepted offer price during winter 2022/23 was approximately £240/MWh, 159% higher than the prices registered during winter 2020/21 (LCP Delta, 2023).

The DFS was an enhanced action that was available to NGESO's control room to tender contracts to reduce demand during system scarcity. Partly because of its status as a demonstrator project, NGESO offered a Guaranteed

Acceptance Price of £3,000/MWh for test events. There was no Guaranteed Acceptance Price for live events, but prices exceeded the test event Guaranteed Acceptance Price, ranging from £4,400 to £6,400/MWh (LCP Delta, 2023, National Grid, 2023*i*). These prices were more than 10 times higher than the Balancing Mechanism's capacity weighted prices during winter 2022-2023, as we show in Table AT.30.

However, this comparison may be unfair to the DFS. The appropriate comparison may be to other enhanced services that NGESO procured, including back-up services that were rarely utilized. The use of the remaining five coal power plants in Great Britain was an important variable in considering costs of enhanced services.[47] LCPDelta's Balancing Costs Review, conducted on behalf of National Grid NGESO, wrote the following regarding "changes in market behaviours":

> *Last winter, ≈2GW of coal capacity consistently offered its generation into the Balancing Mechanism at prices close to £4,000/MWh. This meant the units ahead of coal in the merit order, such as inflexible combined cycle gas turbine units, were able to price their offers close to the £4,000/MWh with the same likelihood of being accepted as they had before. This led to very extreme Balancing Mechanism prices and costs on some days. This winter the coal units were given coal contingency contracts and operated outside of the Balancing Mechanism. This removed the upper limit of £4,000/MWh for units to price up to and led to average peak prices dropping. The removal of this upper bound did however lead to some rare occasions when even more extreme prices were seen, as the marginal units were able to set their own ceiling, and on the 12th December prices as high as £6,000/MWh were accepted.*

These coal contingency contracts were only utilized once, on March 07, 2023. Somewhat puzzlingly, that day's total balancing costs (£4.4M) and volume of actions (20 GWh) were lower than other key days during the winter. For example, on December 12, 2022, balancing costs reached £27.2M for 23 GWh of actions; and on December 29, 2022, balancing costs reached £11.9M for 49 GWh of actions (LCP Delta, 2023).

The coal power plants on these contingency contracts were instructed to warm up by NGESO in another six instances over the winter: 12 December, 2022; January 23, 2023; January 24, 2023; January 26, 2023; February 7, 2023; and February 8, 2023. On these days, NGESO paid approximately £6,000 per hour of time the plants were "warm" to synchronize them with the grid frequency, even though they were not utilized (Horgan, 2023*a*).

NGESO explained that they signed these winter contingency contracts "to ensure safe and secure operation of the electricity system throughout Winter" (National Grid, 2022*d*). With a total capacity of 2.2 GW procured, these contracts cost NGESO £340M to £395M (John, 2023, National Grid, 2022*d*). Only two units on the West Burton A power plants were used on March 7, 2023 during a total of seven hours, with a total volume delivered of 2.5 GWh.

By comparison, DFS cost approximately £11.1M ((National Grid, 2023*d*) and calculated from National Grid (2023*g*,*h*)), 2.8% of the capacity payments spent on the contingency coal contracts. Officially, the DFS delivered 3.3GWh of demand reduction, though our difference-in-differences calculations of Octopus Energy's official versus actual demand reduction might lead one to judge 'actual' demand reduction to be closer to 2.9 GWh (assuming an overestimation of ≈13% in the DFS official figures). Viewed in this light, the costs associated with DFS were arguably small compared to the £340M to £395M procurement costs associated with the contingency coal contracts, despite comparable total utilization.

5.3 Welfare analysis

We have completed a welfare analysis following Finkelstein and Hendren (2020) and Hendren and Sprung-Keyser (2020), calculating the marginal value of public funds (MVPF) associated with DFS. The MVPF compares the

[47] During winter 2022/23, the five coal power units receiving contingency payments were: West Burton A (two units), Drax (two units), and Ratcliffe (one unit). See National Grid (2022*d*).

marginal social benefits of the policy to the net cost to the government. In this case, to simplify our analysis, we consider NGESO to be close enough to the Government that we assume their costs to be the Government's.

$$\text{MVPF} = \frac{\text{Beneficiaries' willingness to pay}}{\text{Net Costs}} \tag{21}$$

We focused on the costs and benefits of Octopus Energy's Saving Sessions only. We had individual data for Octopus Energy customers and thus have the ability to estimate their actual demand reduction from our difference-in-differences, in contrast to demand reduction from other DFS providers. We then used these costs and benefits to calculate the MVPF for Octopus Energy's Saving Sessions.

5.3.1 Costs

NGESO paid Octopus Energy £3/kWh of electricity demand reduction during test events, as discussed in Section 2.2. Demand reduction was measured using P376 baselines and clipping demand reduction at 0, as explained in Section 2.4. For live events, prices were settled through a private auction. In Section 2.3, we showed how Octopus Energy rewarded customers with different incentives for their participation in the two live events. We assumed that Octopus Energy retained the same amount of money per customer per kWh of demand reduction – £0.75. We thus obtain assumed auction prices of £4.125 and £4.75. A summary of these assumptions is presented in Table AT.31.

The mechanical costs for NGESO for each individual event are equal to NGESO payments multiplied by total demand reduction as calculated using the P376 methodology and "clipping" demand reduction in any given half-hour at 0, as discussed in Section 2.4.

Fiscal externalities to the DFS intervention are difficult to measure. Do generators or suppliers change their generation schedules or procurement strategies in response to changes in wholesale prices brought on by DFS displacing some marginal generation, thus affecting tax revenue? We think these effects are unlikely or very small too. We have thus ignored fiscal externalities in this analysis. We also ignore effects of demand reduction on Octopus Energy's income and Government revenue. Lower demand reduces customer expenditures, which reduces Octopus Energy income and/or Government tax revenue (through VAT and corporate income taxes). However, these effects will be exactly offset by lower energy bills for customers, and the welfare effects would be zero overall. Due to the envelope theorem, consumers are indifferent between their pre-Savings Session consumption and the Savings Sessions incentive with the change in their consumption, thus we do not add up the energy savings as benefits to the consumers.

Thus the costs considered for this calculation are:

$$\text{Costs} \quad = \quad \text{P376 clipped demand reduction} \; * \; \text{ESO payments [Test/Live] per demand reduction} \tag{22}$$

5.3.2 Benefits

We consider three main benefits of the implementation of the DFS service: revenue to consumers and suppliers, CO_2eq savings, and avoided probability of lost load (blackouts).

Payments to consumers and suppliers. We assume DFS payments to consumers and Octopus Energy are a benefit to these recipients. We assume no costs by consumers (in turning down) nor by Octopus Energy (in setting up the Saving Sessions program). We then calculate:

$$\text{Payments} \quad = \quad \text{P376 clipped demand reduction} \; * \; \text{ESO payments [Test/Live] per demand reduction} \tag{23}$$

CO_2eq abatement benefits. Demand reduction causes societal benefits associated with pollution reduction associated with electricity generation.[48] NGESO introduced DFS as an enhanced action. Enhanced actions take place when NGESO has exhausted other market actions such as buying electricity from continental Europe, reconfiguring combined cycle gas turbine dispatch levels, or calling on cold generators (including some of Great Britain's coal generators) to warm up and be ready for dispatch in the Balancing Mechanism. NGESO published the "order of action" of these services in their "DFS Deep Dives" slides (National Grid, 2022*c*, 2023*e*). The DFS and the Winter Contingency Coal Contracts were conceived as last-resort options to prevent or reduce any requirement for demand disconnection when all other options had been exhausted.

Under these conditions, we assume that the Saving Sessions displaced either 1) coal that would have provided reserve capacity, or 2) the marginal generation in the merit order for the settlement periods during Saving Sessions when Octopus Energy customers reduced their demand. We note the following considerations and assumptions:

- Given the high gas prices during winter 2022-2023, gas combined cycle gas turbines were always the marginal fuel in the Balancing Mechanism merit order as seen in Figure AI.5.
- One might have thought that the distinction between live versus test events under the DFS causes different scenarios for marginal plants. Indeed, during some Saving Sessions, NGESO instructed some coal power plants to warm up (Grid Beyond, 2023). However, this happened not only during the live events on January 23rd and 24th 2023, but also during the test event on Dec 12th 2022. On none of these occasions was the coal actually utilized; the first time that the backup coal generation was used was on March 7, 2023, during a cold snap in Great Britain (Gillespie, 2023).

In summary, we assumed that:

1. During events when NGESO's control room did not instruct back-up coal generators to warm up, the marginal generator was always a combined cycle gas turbine, and Saving Sessions demand reduction displaced this gas generation.

2. When coal generators were instructed to warm up by the control room, Saving Sessions demand reduction displaced their use as an alternative enhanced action. Functionally, this means that we assumed that the marginal fuel displaced was coal on the Saving Sessions of December 12, 2022, January 23, 2023, and January 24, 2023.

In order to account for the CO_2 abatement caused by Saving Sessions, we used the valuation of greenhouse gas emissions for policy appraisal and evaluation published by the UK Government in their central series (UK GOV, 2021): 248 $£_{2020}/tCO_2$ for Sessions held in 2022 and 252 $£_{2020}/tCO_2$ for Sessions held in 2023. These prices are in $£_{2020}$; we adjust for inflation using the Bank of England inflation calculator (Bank of England, 2023). For both gas and coal, we used the direct emissions of specific generation technologies reported by the Intergovernmental Panel on Climate Change of combined cycle gas turbines and coal plants (Bruckner et al., 2014). We calculated these savings based on the demand reduction we measure using our Octopus versus Bulb DiD, as described in

[48]We have not included co-benefits of CO_2 abatement in this analysis. We analyzed additional benefits from air quality improvements associated with abatement of pollutants associated with electricity generation (UK GOV, 2023*a*). The emission factors of NO_x, SO_x and PM2.5 pollutants from Department for Environment, Food and Rural Affairs (Defra) (2023) and the damage costs of these pollutants can be consulted in Table AT.32. The contribution of these additional benefits is small when compared to the CO_2eq savings that we obtain in Table 12, only accounting for 0.4% or 7.4% of these savings when the marginal fuel is gas or coal respectively. Due to these co-benefits' small overall contribution to the total benefits calculation (2.2% of the total CO_2eq savings), we ignore them in this main analysis. However, it is important to highlight that the literature does not present a comprehensive framework to quantify these co-benefits outside of air quality and impacts on health (which have regional and local effects), and their value could be higher than currently estimated (Brockway and Finn, 2022, Jennings et al., 2020).

Section 5.1. In summary, we calculated CO2 savings from Saving Sessions as follows:

$$\text{CO}_2 \text{ Savings} \quad = \quad \text{DiD demand reduction} \, * \, \text{CO}_2 \text{ emissions [gas/coal]} \, * \, \text{social cost of CO}_2 \text{ in £2023} \quad (24)$$

Using the demand reduction calculated through our DiD methodology, as noted in the fourth row in Table AT.30, we show estimated emission savings in Table 12:

Table 12: DiD estimated demand reduction and CO_2eq emission savings per Saving Session.

Saving Session date	DiD demand reduction by Octopus Energy (MWh)	CO_2eq reduction (tCO_2eq)	CO_2eq savings (£)
November 15, 2022	93.02	70.70	21,022.75
November 22, 2022	121.68	92.48	27,500.05
November 30, 2022	111.46	84.71	25,189.81
December 1, 2022	92.83	70.55	20,978.92
December 12, 2022	277.57	102.70	30,538.84
January 19, 2023	68.80	52.29	15,549.04
January 23, 2023	185.99	68.82	20,463.34
January 24, 2023	216.33	80.04	23,801.55
January 30, 2023	67.89	51.60	15,343.83
February 13, 2023	107.45	81.66	24,283.66
February 21, 2023	106.94	81.28	24,168.84
March 15, 2023	113.67	86.39	25,689.78
March 23, 2023	78.42	59.60	17,723.33
Total	1,642.09	982.83	292,253.74

Note: Using the results of our Octopus versus Bulb DiD, specifically the LATE on sign-up, we estimated the demand reduction achieved in each of the Saving Sessions that Octopus Energy implemented throughout Winter 2022-2023. These results differed from the official demand reduction calculated with the P376 methodology suggested by NGESO, with our estimation usually lower than the demand reduction estimated using NGESO's methodology. By assessing which was the marginal generation for each of the half-hours in which the Saving Sessions occurred, we inferred the fuel generation that was avoided – in all instances, either gas or coal. We assumed that if coal plants under the Coal Contingency Contracts were called to warm up, coal generation was displaced, irrespective of the marginal plant in the Balancing Mechanism providing services. With the corresponding CO_2 emissions for each of these fuels and using the calculated demand reduction from the DiD, we then calculated the amount of CO_2eq emissions avoided.

Other benefits: Value of Lost Load. In extreme scenarios and/or in the future, services like the DFS will not only displace gas and coal generation, but will also alleviate insufficient reserve capacity where there is a danger of lost load (blackouts). Lost loads may increase in probability as renewable penetration of the UK grid increases; in addition, coal power plants, including all three of the coal power plants that were in service during the winter of 2022-2023, are steadily being retired over the next two years.[49]

The value of lost load (VoLL), which is currently set at £6,000/MWh by the Balancing and Settlement Code (National Grid, 2022*a*), signals the value to domestic and non-domestic consumers in Great Britain of having secure and continuous electricity supply. This estimate is in a similar range to the VoLL of $4,300/MWh found by Brown and Muehlenbachs (2023) based on quasi-experimental evidence based on battery adoption among customers of California's largest electric utility, Pacific Gas and Electric, though much higher than the $1,500/MWh found by

[49] After winter 2023-24, NGESO began talks with EDF and Drax about their coal units in West Burton and Yorkshire. Some of these units provided contingency services during winter 2022-23 through the coal contingency contracts, while other units at these sites sold electricity in wholesale markets and the balancing mechanisms. Both companies have already confirmed they were unable to strike a deal with NGESO to continue operations past winter 2022-23 (Drax, 2023, EDF, 2023). Uniper's Ratcliffe-on-Soar has confirmed that the plant will stay online but retire in September 2024 (Horgan, 2023*b*).

Khanna and Rowe (2023) based on quasi-experimental evidence in Delhi.[50] With this in mind, we have calculated a second MVPF where we assume that demand reduction caused by Saving Sessions is valued at the administrative VoLL (£6,000/MWh).

$$\text{VoLL} = \text{DiD demand reduction} * £6{,}000/\text{MWh} \tag{25}$$

5.4 Calculation of MVPF for each Saving Session

Using the assumptions outlined above, we calculated the MVPF for each individual Saving Session within the DFS, presented in Table 13.[51] We show the inputs and outputs of the MVPF visually in Figure 11. The results suggested that the DFS was most valuable when used to avoid cases in which there is a high chance of lost load in the absence of compensatory actions.

Table 13: Values of MVPF for each individual Saving Session and for the whole program.

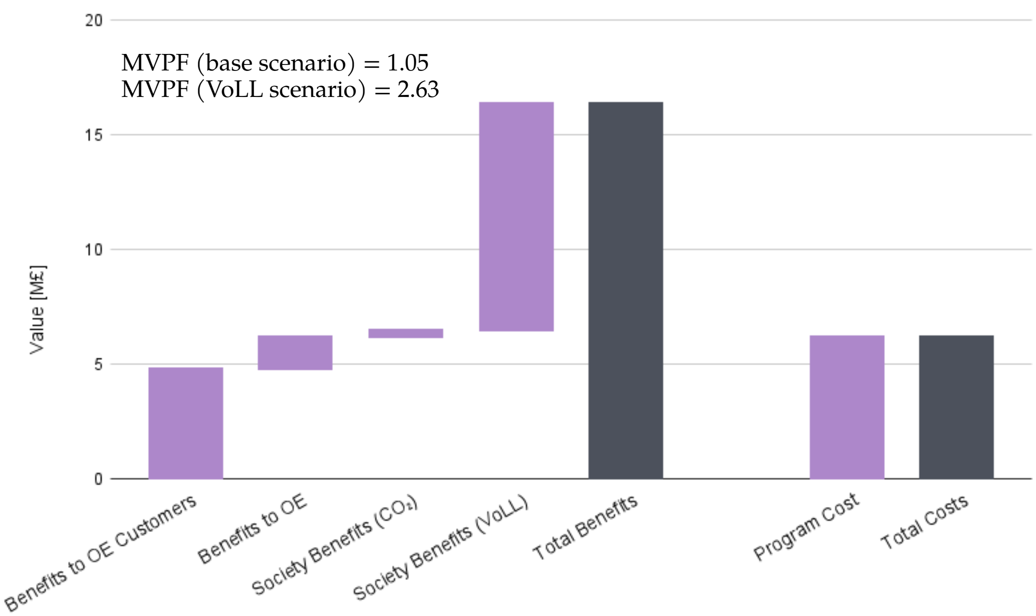
Day	MVPF	MVPF (VoLL scenario)
November 15, 2022	1.06	2.76
November 22, 2022	1.07	3.29
November 30, 2022	1.06	3.01
December 1, 2022	1.06	2.78
December 12, 2022	1.02	3.03
January 19, 2023	1.05	2.74
January 23, 2023	1.02	2.47
January 24, 2023	1.02	2.12
January 2023	1.05	2.42
February 13, 2023	1.05	2.63
February 21, 2023	1.05	2.44
March 15, 2023	1.05	3.13
March 23, 2023	1.04	2.31
Weighted average	**1.05**	**2.63**

Note: The MVPF of Saving Sessions we calculated using the costs and benefits outlined in Sections 5.3.1 and 5.3.2. The MVPF ignoring the Value of Lost Load (VoLL) suggested positive but limited additional value from DFS, as the MVPF calculated is very close to 1. However, where we assume that the DFS avoids the use of actions valued at the level of VoLL, as explained in Section 5.3.2, the MVPF rises to values in between 2.12 and 3.29 depending on the Saving Sessions studied.

[50] Interestingly, the "administrative VoLL value" of £6,000/MWh may be lower than the true VoLL from NGESO's perspective, as revealed by the fact that NGESO has taken some actions whose cost was above £6,000/MWh (National Grid, 2022*a*). Other Balancing Mechanism prices have reached levels close to £6,000/MWh, for instance during the Saving Session run on 12 December, 2022 (National Grid, 2023*e*). Relatedly, the BSC's £6,000/MWh VoLL estimate is based on a report commissioned by Ofgem and prepared by London Economics in 2013, which set a weighted average VoLL across domestic and SME users for the winter peak weekday figures of £16,940/MWh (Economics, 2013).

[51] In these calculations, as outlined in Sections 5.3.1 and 5.3.2 we use the demand reduction derived from our Octopus versus Bulb DiD to value the CO_2 carbon emissions abatement and value of lost load. If we used the P376 methodology, which over-estimates this demand reduction by a factor of 13% as seen in Table AT.30, we would obtain slightly higher MVPF values. These can be see in Table AT.33 and compared to the ones in Table 13.

Figure 11: Summary of costs and benefits for the whole Saving Session campaign.



Note: We calculated benefits and costs benefits as outlined in Sections 5.3.1 and 5.3.2 Note that we calculated the benefits to OE Customers and OE, and the costs to NGESO, using the P376 baseline methodology (see Table AT.30), which we believe overestimated demand response by 13%, as DFS providers and NGESO used used this methodology to settle flexibility delivered. In contrast, we calculated Social Benefits using "real" demand reduction as estimated through our DiD methodology (see Table AT.30). As explained in Section 5.3.1, we have ignored fiscal externalities and additional health benefits from GHG emissions in this analysis as they are likely to be very small.

6 Conclusion

As electricity grids around the world increase their share of generation from renewable but non-dispatchable sources, the importance of demand side response to grid stability will increase (Lever et al., 2021, Mata et al., 2020, National Grid, 2023*j*, Sanders et al., 2016). In the past, industrial and commercial consumers have provided most demand side response (Element Energy, 2012, Warren, 2014). Domestic consumers are an important source of further response.

We analyzed the UK's largest ever demand response program to measure its impact on energy demand and economic welfare. To understand the impact of this program, the UK National Grid Electricity System Operator (NGESO) used a specific methodology for calculating demand reduction, governed by the P376 amendment to Great Britain's electricity balancing and settlement code. This technique is akin to a "pre-post" comparison whereby customers' consumption before a Saving Session is used as their "baseline" from which to calculate demand reduction, defined as the difference between the result of this baseline and their actual consumption. Much of the impetus for our research stemmed from our concern about potential biases associated with this approach, such as selection.

Using our Octopus versus Bulb DiD, we found that simply inviting customers to sign up to Saving Sessions is associated with a ≈10% reduction in consumption during Saving Sessions. Using all three of our DiDs, we found that signing up to participate in DFS events reduced demand by ≈25% during Saving Sessions. Additionally, we found that "opting in" to participate in Saving Sessions reduced demand by ≈40% during the campaign.

These effect were larger than those usually seen in large-sample randomized control trials. For instance, the impact of home energy reports has been estimated to be a 1-3% reduction in customers' energy consumption (Allcott and Rogers (2014), Brandon et al. (2017)). However, the effects we identify are in line with the impact of time-of-use tariffs and other technologies to reduce demand for specific hours of the day, usually identified in much smaller-sample settings (Crawley et al., 2021, Mata et al., 2020, Stromback et al., 2011). As with these analyses of efforts to reduce consumption during key peak times, Saving Sessions involved people reducing demand for just 29 half-hours during the winter period.

We found differences between our estimated demand reduction and those estimated using the methodology endorsed by NGESO. Across Octopus Energy's customer base during those half-hours, the demand reduction totalled 1642 MWh demand reduction, over 14.5 hours. This estimate is approximately 13% lower than the official 1860 MWh demand reduction as measured by Octopus Energy following NGESO's "pre-post" methodology. We believe that this suggests an upward bias in the NGESO methodology, which comes from both measurement error and selection. Still, there is a relatively close concordance between results obtained from our DiDs and those obtained using NGESO's preferred methodology for each individual event.

We also examined how changing notice period and incentive level (£ per kWh demand reduction) changes customers' response. Almost all Saving Sessions featured day-ahead notice, and customers tended to receive this day-ahead notice around the same time for any given Session. Moreover, although the incentive level somewhat varied between Sessions, it never varied between customers within a Session.

Two Saving Sessions (i.e., 13 Feb 2023 and 15 March 2023) featured different notice periods and financial incentives provided to customers. Analyzing data from the 13 Feb 2023 Session, we found that intraday (instead of day-ahead) notice increased in-Session consumption by $\approx$7.1%. This extra consumption represents approximately 25% of customers' Saving Sessions demand response based on our DiDs. The 15 March 2023 Session featured intraday notice for most customers with a randomized subset receiving a special "heads-up" email. This supplementary messaging is arguably "softer" than the standard day-ahead notices, as customers could not use the heads-up email to opt into the Saving Session. And while the email said there may be a Session the next day, this was not guaranteed. Despite being non-committal, we found that the day-ahead heads-up email further decreased in-Session consumption by 2-3%, approximately 7% of typical demand reduction achieved by signed up customers in that Session. In terms of Session participation, we also found higher opt-in rates when customers have longer notice. We found higher opt-in and lower consumption among a random subset of customers who received an SMS with an extra £1.25 incentive if they reduced their demand. However, the consumption differences were not precise enough to rule out the possibility of no difference between these customers and the "business-as-usual" customers who did not receive this SMS.

Overall, our welfare analysis suggested that the program yielded positive benefits relative to the costs involved. Specifically, the MVPF when ignoring the value of lost load was 1.05, with only very small marginal benefits. However, when we ascribed the UK's official value of lost load to each MWh of demand reduction, we found a large MVPF of DFS of 2.6. In other words, DFS was more valuable in cases in which there was a high chance of lost load, which may increase in probability as renewable penetration increases and coal power plants are retired in Great Britain over the next two years. In addition, if the high Guaranteed Acceptance Price (£3,000 per MWh in the 2022-23 Saving Sessions) becomes lower or unnecessary over time, the MVPF is likely to increase.

Our welfare analysis made important simplifying assumptions. In particular, we assumed that it was cost-less for both 1) DFS providers like Octopus Energy to deliver DFS implementations like Saving Sessions, and 2) for customers to deliver demand reduction. Further research where the incentive per kWh demand reduction varies between customers will be crucial to elucidate the consumer welfare costs of delivering flexibility.

Finally, our analyses suggest a tension between the value and magnitude of flexibility response. We believe it is reasonable to assume that grid operators such as NGESO find it more difficult to forecast lost load in a given half-hour for half-hours further in the future. If it is correct that lost load becomes more certain when the half-hour in question approaches, our results from our RDD and field experiment suggest that grid operators and policymakers face a potential trade-off. If the notice period required from NGESO and grid operators is shorter, the flexibility is more valuable. Yet, our results show that domestic customers' flexibility response is smaller, though still substantial, when the notice they receive is closer to the time of flexibility "delivery".

References

Al-Ubaydli, O., Lai, C.-Y. and List, J. A. (2023), A simple rational expectations model of the voltage effect, Technical report, National Bureau of Economic Research. 3

Allcott, H., Collard-Wexler, A. and O'Connell, S. D. (2016), 'How do electricity shortages affect industry? evidence from india', *American Economic Review* **106**(3), 587–624. 2

Allcott, H. and Rogers, T. (2014), 'The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation', *American Economic Review* **104**(10), 3003–3037. 52

Angrist, J. D. (2006), 'Instrumental Variables Methods in Experimental Criminological Research: What, Why and How?', *Journal of Experimental Criminology* **2**(1), 23–44. 40, 79

Angrist, J. D. and Pischke, J.-S. (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press. 79

Baker, A. C., Larcker, D. F. and Wang, C. C. Y. (2022), 'How much should we trust staggered difference-in-differences estimates?', *Journal of Financial Economics* **144**(2), 370–395. 14

Baldick, R., Kolos, S. and Tompaidis, S. (2006), 'Interruptible electricity contracts from an electricity retailer's point of view: Valuation and optimal interruption', *Operations Research* **54**(4), 627–642. 2

Bank of England (2023), 'Inflation calculator'.
URL: *https://www.bankofengland.co.uk/monetary-policy/inflation/inflation-calculator* 48

Barreca, A. I., Guldi, M., Lindo, J. M. and Waddell, G. R. (2011), 'Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification', *The Quarterly Journal of Economics* **126**(4), 2117–2123. 33, 72

Barreca, A. I., Lindo, J. M. and Waddell, G. R. (2016), 'Heaping-Induced Bias in Regression-Discontinuity Designs', *Economic Inquiry* **54**(1), 268–293. 33, 72

Basu, A., Coe, N. B. and Chapman, C. G. (2018), '2SLS Versus 2SRI: Appropriate Methods for Rare Outcomes and/or Rare Exposures', *Health Economics* **27**(6), 937–955. 74, 79

Battey, H. S., Cox, D. R. and Jackson, M. V. (2019), 'On the Linear in Probability Model for Binary Data', *Royal Society Open Science* **6**(5), 190067. 79

Bergquist, M., Thiel, M., Goldberg, M. H. and Van Der Linden, S. (2023), 'Field Interventions for Climate Change Mitigation Behaviors: A Second-Order Meta-Analysis', *Proceedings of the National Academy of Sciences* **120**(13), e2214851120. 2, 3, 11

Bertrand, M., Duflo, E. and Mullainathan, S. (2004), 'How much should we trust differences-in-differences estimates?*', *The Quarterly Journal of Economics* **119**(1), 249–275. 15, 22

Bhattacharya, J., Goldman, D. and McCaffrey, D. (2006), 'Estimating Probit Models with Self-Selected Treatments', *Statistics in Medicine* **25**(3), 389–413. 79

Bhuller, M. and Sigstad, H. (2022), '2SLS with Multiple Treatments'. 75

Brandon, A., Ferraro, P. J., List, J. A., Metcalfe, R. D., Price, M. K. and Rundhammer, F. (2017), 'Do the effects of nudges persist? theory and evidence from 38 natural field experiments', (23277). DOI: 10.3386/w23277.
URL: *https://www.nber.org/papers/w23277* 52

Brandon, A., List, J. A., Metcalfe, R. D., Price, M. K. and 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. 2

Breen, R., Karlson, K. B. and Holm, A. (2018), 'Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models', *Annual Review of Sociology* **44**(1), 39–54. 79

Brewer, D. and Crozier, J. (2023), 'Who heeds the call to conserve in an energy emergency? evidence from smart thermostat data', *Evidence from Smart Thermostat Data (January 29, 2023)* . 3

Brockway, P. and Finn, O. (2022), 'Co-benefits of energy demand reduction are much broader than health'.
URL: *https://www.creds.ac.uk/co-benefits-of-energy-demand-reduction-are-much-broader-than-health/* 48

Brown, D. P. and Muehlenbachs, L. A. (2023), *The Value of Electricity Reliability: Evidence from Battery Adoption*, University of Alberta, Faculty of Arts, Department of Economics. 49

Bruckner, T., Schlömer, S., Hänsel, G., de Jager, D. and Neelis, M. (2014), 'Technology-specific cost and performance parameters', *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* .
URL: *https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_annex-iii.pdfpage=7* 48

Burgess, R., Greenstone, M., Ryan, N. and Sudarshan, A. (2020), 'The consequences of treating electricity as a right', *Journal of Economic Perspectives* **34**(1), 145–169. 2

Calonico, S., Cattaneo, M. D., Farrell, M. H. and Titiunik, R. (2017), 'Rdrobust: Software for Regression-Discontinuity Designs', *The Stata Journal: Promoting communications on statistics and Stata* **17**(2), 372–404. 72

Carreras, B. A., Colet, P., Reynolds-Barredo, J. M. and Gomila, D. (2021), 'Assessing blackout risk with high penetration of variable renewable energies', *IEEE Access* **9**, 132663–132674. 2

Cattaneo, M. D., Idrobo, N. and Titiunik, R. (2019), *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cambridge University Press. 72

Cattaneo, M. D., Idrobo, N. and Titiunik, R. (Forthcoming), *A Practical Introduction to Regression Discontinuity Designs: Extensions*, Cambridge Elements: Quantitative and Computational Methods for Social Science, Cambridge University Press. 33, 36, 72

Cattaneo, M. D. and Vazquez-Bare, G. (2017), 'The Choice of Neighborhood in Regression Discontinuity Designs', *Observational Studies* **3**(2), 134–146. 72, 73

Caves, D. W. and Christensen, L. R. (1980), 'Econometric analysis of residential time-of-use electricity pricing experiments', *Journal of Econometrics* **14**(3), 287–306. 2

Caves, D. W., Christensen, L. R. and Herriges, J. A. (1984), 'Consistency of residential customer response in time-of-use electricity pricing experiments', *Journal of Econometrics* **26**(1-2), 179–203. 2

Chiburis, R. C., Das, J. and Lokshin, M. (2012), 'A Practical Comparison of the Bivariate Probit and Linear IV Estimators', *Economics Letters* **117**(3), 762–766. 79

Cole, M. A., Elliott, R. J., Occhiali, G. and Strobl, E. (2018), 'Power outages and firm performance in sub-saharan africa', *Journal of Development Economics* **134**, 150–159. 2

Costa, F. and Gerard, F. (2021), 'Hysteresis and the welfare effect of corrective policies: Theory and evidence from an energy-saving program', *Journal of Political Economy* **129**(6), 1705–1743. 3

Crawley, J., Johnson, C., Calver, P. and Fell, M. (2021), 'Demand response beyond the numbers: A critical reappraisal of flexibility in two united kingdom field trials', *Energy Research & Social Science* **75**, 102032. 52

Cunningham, S. (2021), *Causal Inference: The Mixtape*, Yale University Press. 74

Department for Environment, Food and Rural Affairs (Defra) (2023), 'Emission factors - defra, uk'.
URL: *https://naei.beis.gov.uk/data/emission-factors* 48, 122

Dong, Y. and Lewbel, A. (2015), 'A Simple Estimator for Binary Choice Models with Endogenous Regressors', *Econometric Reviews* **34**(1-2), 82–105. 79

Drax (2023), 'Drax ends half a century of coal-fired power generation - drax global'.
URL: *https://www.drax.com/press_release/drax-ends-half-a-century-of-coal-fired-power-generation/* 49

Economics, L. (2013), 'The value of lost load (voll) for electricity in great britain', *Final report for OFGEM and DECC* . 50

EDF (2023), 'End of generation at west burton a'.
URL: *https://www.edfenergy.com/media-centre/news-releases/end-generation-west-burton* 49

Element Energy (2012), 'Demand side response in the non-domestic sector. Final report for Ofgem', p. 69. 51

Elexon BSC (2023), 'P376 "utilising a baselining methodology to set physical notifications"'.
URL: *https://www.elexon.co.uk/mod-proposal/p376/* 12

Farronato, C., Fong, J. and Fradkin, A. (2023), 'Dog eat dog: Balancing network effects and differentiation in a digital platform merger', *Management Science* . 4

Feng, K., Ouyang, M. and Lin, N. (2022), 'Tropical cyclone-blackout-heatwave compound hazard resilience in a changing climate', *Nature Communications* **13**(11), 4421. 2

Finkelstein, A. and Hendren, N. (2020), 'Welfare analysis meets causal inference', *Journal of Economic Perspectives* **34**(4), 146–167. 46

Fisher-Vanden, K., Mansur, E. T. and Wang, Q. J. (2015), 'Electricity shortages and firm productivity: evidence from china's industrial firms', *Journal of Development Economics* **114**, 172–188. 2

Fotis, G., Vita, V. and Maris, T. I. (2023), 'Risks in the european transmission system and a novel restoration strategy for a power system after a major blackout', *Applied Sciences* **13**(11), 83. 2

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M. and Gelman, A. (2019), 'Visualization in Bayesian Workflow', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **182**(2), 389–402. 79

Gelman, A., Hill, J. and Vehtari, A. (2020), *Regression and Other Stories*, 1 edn, Cambridge University Press. 11, 13, 17, 19, 28, 32, 33, 36, 40, 41, 73, 74, 75, 79

Gelman, A. and Imbens, G. (2019), 'Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs', *Journal of Business & Economic Statistics* **37**(3), 447–456. 72

Gelman, A. and Zelizer, A. (2015), 'Evidence on the Deleterious Impact of Sustained use of Polynomial Regression on Causal Inference', *Research & Politics* **2**(1), 205316801556983. 72

Gertler, P. J., Shelef, O., Wolfram, C. D. and Fuchs, A. (2016), 'The demand for energy-using assets among the world's rising middle classes', *American Economic Review* **106**(6), 1366–1401. 2

Gillespie, T. (2023), 'National grid uses back-up coal plant for first time to secure uk power supply'.
URL: *https://www.bloomberg.com/news/articles/2023-03-07/uk-grid-uses-back-up-coal-plant-for-first-time-to-secure-supply* 48

Gomila, R. (2021), 'Logistic or Linear? Estimating Causal Effects of Experimental Treatments on Binary Outcomes Using Regression Analysis.', *Journal of Experimental Psychology: General* **150**(4), 700–709. 79

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* **225**(2), 254–277. 14

Greene, W. (2019), *Econometric Analysis*, 8th edition edn, Pearson. 19, 40

Grid Beyond (2023), 'Eso issues first electricity margin notice of 2023 and calls contingency coal units'.
URL: *https://gridbeyond.com/eso-issues-first-electricity-margin-notice-of-2023-and-calls-contingency-coal-units/* 48

Grimwood, T. (2023), ''power rinse': Households game demand flexibility service'.
URL: *https://utilityweek.co.uk/power-rinse-households-game-demand-flexibility-service/* 13

Grubb, M., Ferguson, T., Musat, A., Maximov, S., Zhang, Z., Price, J. and Drummond, P. (2022), 'Navigating the crises in european energy: Price inflation, marginal cost pricing, and principles for electricity market redesign in an era of low-carbon transition'.
URL: *https://www.ineteconomics.org/research/research-papers/navigating-the-crises-in-european-energy* 7

Harvey, L. A. (2018), 'Statistical Testing for Baseline Differences Between Randomised Groups is not Meaningful', *Spinal Cord* **56**(10), 919–919. 75

Hendren, N. and Sprung-Keyser, B. (2020), 'A unified welfare analysis of government policies*', *The Quarterly Journal of Economics* **135**(3), 1209–1318. Citation Key: 10.1093/qje/qjaa006 tex.eprint: https://academic.oup.com/qje/article-pdf/135/3/1209/33427738/qjaa006.pdf. 6, 46

Hernán, M. M. and Robins, J. M. (2023), *Causal Inference: What If*, Chapman & Hall/CRC. 74

Holladay, J. S., Price, M. K. and Wanamaker, M. (2015), 'The perverse impact of calling for energy conservation', *Journal of Economic Behavior & Organization* **110**, 1–18. 3

Hollenbach, F. M., Montgomery, J. M. and Crespo-Tenorio, A. (2019), 'Bayesian Versus Maximum Likelihood Estimation of Treatment Effects in Bivariate Probit Instrumental Variable Models', *Political Science Research and Methods* **7**(3), 651–659. 79

Horgan, R. (2023*a*), 'Dfs must not punish those unable to take part'.
URL: *https://utilityweek.co.uk/dfs-must-not-punish-those-unable-to-take-part/* 46

Horgan, R. (2023*b*), 'Drax and eso fail to strike deal on coal units returning'.
URL: *https://utilityweek.co.uk/drax-and-eso-fail-to-strike-deal-on-coal-units-returning/* 49

Houthakker, H. S. (1951), 'Electricity tariffs in theory and practice', *Economic Journal* **61**(241), 1–25. 2

Huntington-Klein, N. (2021), *The Effect: An Introduction to Research Design and Causality*, CRC Press, Taylor & Francis Group, Boca Raton. 72, 79

Imai, K., King, G. and Stuart, E. A. (2008), 'Misunderstandings Between Experimentalists and Observationalists about Causal Inference', *Journal of the Royal Statistical Society Series A: Statistics in Society* **171**(2), 481–502. 75, 76

Institute for government (2022), 'Electricity market'.
URL: *https://www.instituteforgovernment.org.uk/article/explainer/electricity-market* 7

Ito, K., Ida, T. and Tanaka, M. (2018), 'Moral suasion and economic incentives: Field experimental evidence from energy demand', *American Economic Journal: Economic Policy* **10**(1), 240–267. 2

Jahn, W., Urban, J. L. and Rein, G. (2022), 'Powerlines and wildfires: Overview, perspectives, and climate change: Could there be more electricity blackouts in the future?', *IEEE Power and Energy Magazine* **20**(1), 16–27. 2

Jennings, N., Fecht, D. and Matteis, S. D. (2020), 'Mapping the co-benefits of climate change action to issues of public concern in the uk: a narrative review', *The Lancet Planetary Health* **4**(9), e424–e433. 48

John, A. (2023), 'Eso puts back-up coal plants to work for first time this winter'.
URL: *https://utilityweek.co.uk/eso-puts-back-up-coal-plants-to-work-for-first-time-this-winter/* 46

Joskow, P. L. (1976), 'Contributions to the theory of marginal cost pricing', *The Bell Journal of Economics* pp. 197–206. 2

Khanna, S. and Rowe, K. (2023), The long-run value of electricity reliability in india. Working paper.
URL: *https://www.imperial.ac.uk/people/s.khanna/document/10882/* 50

Kleibergen, F. and Zivot, E. (2003), 'Bayesian and Classical Approaches to Instrumental Variable Regression', *Journal of Econometrics* **114**(1), 29–72. 79

LCP Delta (2023), 'Balancing costs review'.
URL: *https://www.nationalgrideso.com/document/281781/download* 9, 45, 46, 121

Lee, D. S. and Lemieux, T. (2010), 'Regression Discontinuity Designs in Economics', *Journal of Economic Literature* **48**(2), 281–355. 72

Lever, A., Evans, H., Ravishankar, M., Romanidis, N., Richards, O., Buchanan, F., Pudjianto, D., Strbac, G. et al. (2021), 'Flexibility in great britain'.
URL: *https://publications.carbontrust.com/flex-gb/report* 51

Li, C., Poskitt, D. S., Windmeijer, F. and Zhao, X. (2022), 'Binary Outcomes, OLS, 2SLS and IV Probit', *Econometric Reviews* **41**(8), 859–876. 79

Li, Z. and Agarwal, A. (2017), 'Platform integration and demand spillovers in complementary markets: Evidence from facebook's integration of instagram', *Management Science* **63**(10), 3438–3458. 4

List, J. A. (2022), *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*, Currency. 3

Long, J. S. and Ervin, L. H. (2000), 'Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model', *The American Statistician* **54**(3), 217. 35, 41, 44, 79, 117, 118, 119, 120

Masood, N. A., Yan, R. and Kumar Saha, T. (2018), Cascading contingencies in a renewable dominated power system: Risk of blackouts and its mitigation, *in* '2018 IEEE Power & Energy Society General Meeting (PESGM)', p. 1–5. 2

Mata, , Ottosson, J. and Nilsson, J. (2020), 'A review of flexibility of residential electricity demand as climate solution in four eu countries', *Environmental Research Letters* **15**(7), 073001. 51, 52

Moll, B., Schularick, M. and Zachmann, G. (2023), 'The power of substitution: The great german gas debate in retrospect', *Brookings Papers on Economic Activity* . 2

National Grid (2022*a*), 'Balancing actions above the value of lost load (voll)'.
URL: *https://www.nationalgrideso.com/document/268121/download* 49, 50

National Grid (2022*b*), 'Demand flexibility service guaranteed acceptance price - winter 2022/23'.
URL: *https://www.nationalgrideso.com/document/268856/download* 8

National Grid (2022*c*), 'Order of action - winter 2022'.
URL: *https://www.nationalgrideso.com/document/268116/download* 48

National Grid (2022*d*), 'Winter contingency contracts'.
URL: *https://www.nationalgrideso.com/document/268126/download* 46

National Grid (2023*a*), 'Day ahead constraint flows and limits'.
URL: *https://data.nationalgrideso.com/constraint-management/day-ahead-constraint-flows-and-limits* 80

National Grid (2023*b*), 'Demand flexibility service participation guidance document'.
URL: *https://www.nationalgrideso.com/document/270361/download* 8

National Grid (2023*c*), 'Demand flexibility service procurement rules'.
URL: *https://www.nationalgrideso.com/document/281736/download* 13

National Grid (2023*d*), 'Demand flexibility service. winter 2022/23 review', p. 20.
URL: *https://www.nationalgrideso.com/document/287006/download* 2, 46

National Grid (2023*e*), 'Dfs deep dives'.
URL: *https://www.nationalgrideso.com/document/279196/download* 12, 48, 50

National Grid (2023*f*), 'Electricity markets explained'.
URL: *https://www.nationalgrideso.com/what-we-do/electricity-markets/electricity-markets-explained* 6

National Grid (2023*g*), 'Eso data portal: Demand flexibility service: Live events - dataset'.
URL: *https://data.nationalgrideso.com/dfs/demand-flexibility-service-live-events* 46, 121

National Grid (2023*h*), 'Eso data portal: Demand flexibility service: Test events - dataset'.
URL: *https://data.nationalgrideso.com/dfs/demand-flexibility-service-test-events* 46, 121

National Grid (2023*i*), 'The eso's demand flexibility service'.
URL: *https://www.nationalgrideso.com/electricity-explained/electricity-and-me/esos-demand-flexibility-service* 8, 46

National Grid (2023*j*), 'Future energy scenarios 2023'.
URL: *https://www.nationalgrideso.com/document/283101/download* 51

National Grid (2023*k*), 'Household engagement with the demand flexibility service 20023/23'.
URL: *https://www.nationalgrideso.com/document/283041/download* 13

Office for National Statistics (2022), 'Monthly populations by index of multiple deprivation (imd) decile, england: January 2019 to august 2022'.
URL: *https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/adhocs/15363monthly*
61

Ofgem (2022*a*), 'Decision on the demand flexibility service in relation to an update to the terms and conditions related to balancing'.
URL: *https://www.nationalgrideso.com/document/270366/download* 9

Ofgem (2022*b*), 'Net zero britain: developing an energy system fit for the future'.
URL: *https://www.ofgem.gov.uk/publications/net-zero-britain-developing-energy-system-fit-future* 7

Ofgem (2023*a*), 'Decision on market-wide half-hourly settlement (mhhs) change request cr022'.
URL: *https://www.ofgem.gov.uk/sites/default/files/2023-06/Decision%20on%20MHHS%20Change%20Request%20CR22.pdf* 8

Ofgem (2023*b*), 'Renewable energy guarantees of origin (rego)'.
URL: *https://www.ofgem.gov.uk/environmental-and-social-schemes/renewable-energy-guarantees-origin-rego* 14

Ofgem (2023*c*), 'Retail market indicators'.
URL: *https://www.ofgem.gov.uk/retail-market-indicators* 2

Olexsak, S. J. and Meier, A. (2014), 'The electricity impacts of earth hour: An international comparative analysis of energy-saving behavior', *Energy Research & Social Science* **2**, 159–182. 3

Panteli, M. and Mancarella, P. (2015), 'Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies', *Electric Power Systems Research* **127**, 259–270. 2

Prest, B. C. (2020), 'Peaking interest: How awareness drives the effectiveness of time-of-use electricity pricing', *Journal of the Association of Environmental and Resource Economists* **7**(1), 103–143. 3

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

Rencher, A. C. and Schaalje, G. B. (2008), *Linear Models in Statistics*, 2nd ed edn, Wiley-Interscience, Hoboken, N.J. 79

Sanders, D., Hart, A., Ravishankar, M., Brunert, J., Strbac, G., Aunedi, M. and Pudjianto, D. (2016), 'An analysis of electricity system flexibility for great britain', *Carbon Trust/Imperial College: London, UK* . 51

Savelli, I., Hardy, J., Hepburn, C. and Morstyn, T. (2022), 'Putting wind and solar in their place: Internalising congestion and other system-wide costs with enhanced contracts for difference in great britain', *Energy Economics* **113**, 106218. 80

Senn, S. (1994), 'Testing for Baseline Balance in Clinical Trials', *Statistics in Medicine* **13**(17), 1715–1726. 75

Sheppard, K., Ro, J., bot, S., Lewis, B., Clauss, C., Guangyi, Jeff, Yu, J. Q., Wilson, K., Migrator, L., Thrasibule, WilliamRoyNelson, RENE-CORAIL, X. and vikjam (2023), 'bashtage/linearmodels: Release 5.3'.
URL: *https://github.com/bashtage/linearmodels* 19

Stromback, J., Dromacque, C., Yassin, M. H. and VaasaETT, G. (2011), 'The potential of smart meter enabled programs to increase energy and systems efficiency: a mass pilot comparison short name: Empower demand', *Vaasa ETT* . 52

Tan, C.-W. and Varaiya, P. (1993), 'Interruptible electric power service contracts', *Journal of Economic Dynamics and Control* **17**(3), 495–517. 2

UK GOV (2021), 'Valuation of greenhouse gas emissions: for policy appraisal and evaluation'.
URL: *https://www.gov.uk/government/publications/valuing-greenhouse-gas-emissions-in-policy-appraisal/valuation-of-greenhouse-gas-emissions-for-policy-appraisal-and-evaluation* 48

UK GOV (2023*a*), 'Air quality appraisal'.
URL: *https://www.gov.uk/government/publications/assess-the-impact-of-air-quality/air-quality-appraisal-damage-cost-guidance* 48, 122

UK GOV (2023*b*), 'Energy price guarantee'.
URL: *https://www.gov.uk/government/publications/energy-bills-support/energy-bills-support-factsheet-8-september-2022* 3

Warren, P. (2014), 'A review of demand-side management policy in the uk', *Renewable and Sustainable Energy Reviews* **29**, 941–951. Citation Key: WARREN2014941. 51

Wolak, F. A. (2007), 'Residential customer response to real-time pricing: The anaheim critical peak pricing experiment'. 2

Wolak, F. A. (2011), 'Do residential customers respond to hourly prices? evidence from a dynamic pricing experiment', *American Economic Review* **101**(3), 83–87. 2

Wolak, F. A. (2019), 'The role of efficient pricing in enabling a low-carbon electricity sector', *Economics of Energy & Environmental Policy* **8**(2), 29–52. 8

Wolak, F. A. and Patrick, R. H. (2001), 'The impact of market rules and market structure on the price determination process in the england and wales electricity market'. 7

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, 2nd ed edn, MIT Press, Cambridge, Mass. 19, 40, 79

Yan, R., Masood, N.-A., Kumar Saha, T., Bai, F. and Gu, H. (2018), 'The anatomy of the 2016 south australia blackout: A catastrophic event in a high renewable network', *IEEE Transactions on Power Systems* **33**(5), 5374–5388. 2

Yan, Z., Yang, H., Jia, L. and Tan, Y. (2021), 'Asymmetric network effects in the integration of digital platforms', *Available at SSRN 3871279* . 4

Zakeri, B., Staffell, I., Dodds, P., Grubb, M., Ekins, P., Jääskeläinen, J., Cross, S., Helin, K. and Castagneto-Gissey, G. (2022), 'Energy transitions in europe – role of natural gas in electricity prices', *SSRN Electronic Journal* .
URL: *https://www.ssrn.com/abstract=4170906* 7

Złotecka, D. and Sroka, K. (2018), The characteristics and main causes of power system failures basing on the analysis of previous blackouts in the world, *in* '2018 International Interdisciplinary PhD Workshop (IIPhDW)', p. 257–262. 2

Appendix

AI Additional Information

AI.1 Who participated in Saving Sessions?

As discussed in Section 2.3, customers proactively signed up to Saving Sessions. Thus, those who signed up may have been systematically different from those who did not. Once signed up, customers proactively chose to opt in to each Session, in response to a communication from Octopus Energy, on a Session-by-Session basis. Thus even conditional on sign-up, those who opted in may have been systematically different from those who did not.

AI.1.1 Differences between groups on observable characteristics

We looked at differential sign-up first. Our three DiD strategies exploit three sources of differential sign-up in Saving Sessions: 1) not signing up at all, 2) signing up after most Saving Sessions had happened, and 3) not being invited to sign up due to being a Bulb customer. We examined how these groups differed on observable characteristics in Figures AF.5 to AF.7.

First, we examined how groups differed on customers' region. We used three large regions: 1) Scotland, 2) North England and North Wales, and 3) South England, South Wales, and the Midlands.[52] We found somewhat minimal differences between groups in terms of these three larger regional categories, though Bulb customers were somewhat less likely to reside in North England and North Wales, and more likely to reside in Scotland, the Midlands, South England, and South Wales.

Second, we examined whether the customer was on a "smart" tariff. Smart tariffs are special tariffs for customers who have electric vehicles, batteries, heat pumps, and other 'low carbon technologies'. We considered whether a customer was on a smart tariff as a proxy for their engagement with their home's energy consumption. We found large differences in whether the customer was on a smart tariff between sign-ups and non-sign-ups: 9% of Signed Up Early group were on a smart tariff compared to only 6% of Signed Up Late group and only 3% of non-sign-up customers smart-meter customers.[53]

Third, we showed how customers' Energy Performance Certificate (EPC) letter ratings varied by group. In the UK, an EPC is a property's energy efficiency rating from A (most efficient) to G (least efficient) and is valid for 10 years. EPCs are needed whenever a property is built, sold, or rented. This requirement means that properties without EPCs are more likely to be owner-occupied (rather than rented) properties that have not been sold in the previous 10 years. The distributions of grades were similar across groups, though Octopus customers were somewhat more likely to have no EPC than Bulb customers, as are customers from the Signed Up Late group compared to customers from the Signed Up Early group.

Fourth, we showed the index of multiple deprivation (IMD) for the customer's postcode. IMD is a measure of relative deprivation for each postcode of the UK. We show IMD *quintiles*: a postcode can be classified with levels of deprivation in the following groups: "very low", "low", "medium", "high", and "very high".[54] We found higher

[52] We defined a customer's region based on their district network, a common regional identifier in energy analysis in Great Britain. There are two district networks in **Scotland**: 17 (North Scotland) and 18 (South and Central Scotland). There are four district networks in **North England and North Wales**: 13 (North Wales, Merseyside and Cheshire), 15 (North East England), 16 (North West England), and 23 (Yorkshire). The other eight district networks are in the **Midlands, South England, and South Wales**: 10 (East England), 11 (East Midlands), 12 (London), 19 (South East England), (20 (Southern England), 21 (South Wales), 22 (South West England).

[53] Bulb customers could not sign up to any Octopus Energy smart tariffs, and Bulb did not have a similar set of smart tariffs.

[54] See: Office for National Statistics (2022). Customer postcode IMD was "unknown" for new postcodes that do not yet have an IMD – these were fewer than 1% of Octopus Energy customers.

sign-up in lower-deprivation postcodes. The Signed Up Late group also disproportionately came from lower-deprivation postcodes, but not to the extent that the Signed Up Early group did. The Octopus and Bulb groups showed overall similar distributions in terms of IMD.

Finally, we showed how customers in each group differed in terms of their estimated annual electricity consumption (EAC), a data point that British energy suppliers hold for all customers. As we see in Figures AF.8 to AF.10 the distributions of EAC are very similar for each pair of groups in the three DiD analyses.

Finally, we performed a logistic regression (Table AT.21) of sign-up at any point during the Saving Sessions period on a suite of observable characteristics among all customers that Octopus Energy invited to participate. These included the ones discussed above — EAC, binarized into "high" (greater than or equal to 2,900 kWh/year), or "low" (less than 2,900 kWh/year), where 2,900 kWh/year is a commonly used demarcator of median EAC in Great Britain, IMD quintile, EPC letter grade (with a separate category for no EPC), region (using all 14 district networks as separate categories), and whether the customer was on a smart tariff. We also included a suite of observable characteristics for which we have incomplete data from the DOMUS Property Database: occupancy type (whether the household was a single occupant, a couple, or more than two occupants), floor area (where we categorized 0-68 square meters as "low" floor area homes, 68-90 square meters as "medium" floor area homes, and greater than 90 square meters as "high" floor area homes), age of primary bill payer (in categories of 18-24, 25-34, 35-44, 45-54, 55-64, 75+ years old), and whether the home was in an urban or rural area.[55] Sign-up was higher among customers from lower-deprivation postcodes, where the primary bill-payer was "middle-aged" (45-74 years old), where the home had larger floor areas, and among couples.

AI.1.2 Characteristics of customers who opted in to events

Once customers signed up, they opted into Sessions on a Session-by-Session basis, meaning that opt-ins may be systematically different from non-opt-ins. We investigated this issue using a Poisson regression predicting the number of Sessions a customer opted into, using the same customer characteristics as in our logistic regression predicting sign-up. In examining Table AT.22, we found similar patterns in terms of customer characteristics influencing the number of events opted into as we saw in examining customer characteristics' association with sign-up. "Middle-aged" (45-74 year old) primary bill payers opted in to more events. Couples opted in to more events than multi-occupant households and single-adult households. Large floor area households opted in to more events than medium and small floor area households. Customers in lower-deprivation postcodes opted in to more events than customers in higher-deprivation postcodes.

As another way to examine participation, we examined the likelihood of opting in to a Session, conditional on having signed up to the broader Saving Sessions program before the Session in question (Table AT.23). Again, we found broadly similar associations between customer characteristics and likelihood of opting in.

Finally, we examined an important pair of Sessions, which, unlike any other Sessions, happened on consecutive days – January 23, 2023 and January 24, 2023. They were also unusual in that they were considered "live" rather than "test" events and featured higher compensation per kWh of demand reduction (£3.375/kWh and £4/kWh respectively rather than the usual £2.25/kWh). In examining opt-in to the Session on January 24, 2023, we found the same overall pattern of characteristics associated with opt-in as described in the more general regressions above, with higher opt-in among middle-aged and older primary bill-payers, larger sized properties, lower-deprivation postcodes, couples (rather than multi-occupant households), and smart tariff customers. In our logistic regression predicting sign-up on January 24, 2023, in addition to these covariates, we also controlled for opt-in to January 23, 2023.

In theory, opt-in to January 23, 2023 could be negatively associated with opt-in to January 24, 2023 because

[55] We included "unknown" as a category in all variables from the DOMUS dataset.

of fatigue, or positively associated because it captures latent engagement and/or because customers might form habits of participating in Sessions, where initial participation begets further participation. Our regression model (Table AT.24) is not causal, so we cannot say with confidence which effect dominates. However, we found that opt-in to January 23, 2023 was indeed a very strong predictor of opt-in to January 24, 2023.

AI.2 Mechanisms of Energy Reduction During Saving Sessions: Full Results

Here we explore the behaviors customers used to reduce their domestic electricity consumption during Saving Sessions using two sample surveys. Our sampling frame for these surveys was the population of Octopus Energy customers who had signed up to participate in Saving Sessions by two key dates. On March 20, 2023 (i.e., a few days after the penultimate Saving Session on March 15, 2023), we invited 5,000 randomly-chosen signed-up customers to take part in our first survey. In total, 933 of these customers agreed to take part resulting in a response rate of 18.7%. On April 19, 2023 (i.e., a few weeks after the Saving Sessions campaign had ended), we invited 55,000 randomly-chosen sighed-up customers to take part in our second survey. In total, 4,818 of these customers agreed to take part, resulting in a response rate of 8.8%. Customers invited to take part in our first and second surveys did not overlap. Octopus Energy's marketing team distributed the surveys to the randomly-chosen customers. To do so, this team sent emails to the randomly-chosen customers that included hyperlinks to the surveys. The surveys themselves were built using Typeform, a widely used online-survey platform. The surveys did not explicitly solicit consent to participate. However, Octopus Energy's marketing team sends customers surveys periodically, and customers are always free to respond or withhold response without penalty.[56]

The two surveys had slightly different sets of questions. Both asked customers about their general experience during the Saving Sessions they opted into, whether they would be interested in participating in future Sessions, and what steps they had taken to reduce their electricity consumption during the Saving Sessions they opted into. To design the surveys, we collaborated with Octopus Energy's marketing team given their experience designing surveys for customers and we sent the surveys to colleagues for feedback on the degree to which the questions were understandable and unambiguous.

AI.2.1 Representativeness of Survey Respondents

We compared survey respondents to our population of signed-up Saving Sessions customers using five metrics. The first three metrics were the same metrics we used to compare DFS-participating Octopus Energy customers to Octopus Energy smart meter customers in Section AI.1. Recall that these metrics are: (a) whether a customer has a "smart" tariff; (b) a customer's estimated annual electricity consumption (EAC), which we dichotomize into "high" (i.e., EAC $\geq$ 2,900 kWh/year) and "low" (i.e., EAC $<$ 2,900 kWh/year) EAC; and (c) the index of multiple deprivation (IMD), split into quintiles, for the postcode within which a customer resides. Our remaining two metrics proxied engagement with the peak-pricing campaign itself: (a) the number of Saving Sessions a customer opted into; and (b) the cash value of the average number of OctoPoints a customer earns across the Sessions they opted into.

Survey respondents were broadly representative of the typical customer signed-up to Saving Session sign-up in terms of tariff type, EAC, and postcode IMD (Figure AF.17). However, respondents were somewhat unrepresentative with respect to Saving Sessions engagement, with survey respondents showing particularly high engagement with the Saving Sessions campaign.

[56]We speculate that the reason for the lower response rate in the second survey was due to the reduced salience of Saving Sessions when the marketing team sent the emails to customers (a few weeks after the Sessions had ended, rather than in the middle of the Sessions season).

AI.2.2 Survey Findings

Session Satisfaction and Future Participation. Survey results suggested a high degree of enthusiasm about Saving Sessions. Specifically, over 96% of customers were interested or somewhat interested in continuing participating in Saving Sessions in the future with less than 1% of the respondents responding "Uninterested" or "Very uninterested" (Figure AF.18). And almost 90% of customers said they would be interested in participating in two or more Sessions per week if the program were run again "next Winter" (Figure AF.19). Furthermore, 28% expressed an interest in ten Sessions weekly (i.e., more than one per day). However, notable percentages of the respondents preferred only two Sessions per week (16%) or three sessions per week (12%). Overall, 90% of the responses preferred to have more than two sessions per week.

In the future, grid operators are likely to see period of times during which there is excess electricity, especially in areas near sites of renewable power generation. In such a scenario, it can be beneficial for customers to increase their energy consumption in order to relieve the need to use electricity during more energy-constrained scenarios wherein coal and gas would be used to makeup shortfalls. Thus, we note that 80% of survey respondents responded "Yes" to the question "Would you be interested in being paid to use more energy during times when there's too much electricity on the grid?".

Finally, in response to the question "How much notice would you consider a minimum to prepare for a Saving Session?", 20% of survey respondents said they could prepare with zero to one hour or one to four hours of notice. Nevertheless, most customers said they would need four or more hours notice.

Manual Versus Scheduled Flexibility Strategies. To better understand the behavioral mechanisms of energy reduction, we asked survey respondents who opted into Saving Sessions "What best describes how you participated?" and provided them with a series of non-mutually-exclusive responses. Customers could "tick" agreement with as many of the responses as they wished. 75% of respondents indicated that they engaged in manual demand shifting in that they "manually switched off appliances during the Session and used them at other times". A much smaller group (i.e,. 22% of respondents) indicated that they incorporated *scheduled* demand shifting, agreeing to either (or both) of the response options "scheduled my appliances (like the tumble dryer) to come on before the Session" and/or "scheduled my appliances (like the tumble dryer) to come on after the Session".

We observed that these proportions did not vary very much by subgroup: manual methods of demand shifting were more frequent then scheduled techniques amongst all sub-populations we examined. However, survey respondents on smart tariffs and those with higher estimated annual electricity consumption did *more* scheduling of appliances (Figures AF.20 and AF.21). We observed little difference on these measures in relation to customers' postcode-level deprivation (Figure AF.22).

Looking at how these answers varied by level of per-Session average remuneration (Figure AF.26), two patterns emerged. First, lower-remunerated customers, i.e. those who earned less than 50p, accounted for approximately 6% of survey respondents but 30% of Saving Sessions sign-ups. This group said they did almost all behaviors at lower rates than other customers. Second, changes to electric vehicle (EV) charging were associated with higher average remuneration. (n.b., rows in Figures AF.26 and AF.27 do not sum to 100% because responses are not mutually exclusive.)

We also saw limited differences between behaviors undertaken by customers exhibiting low versus medium versus high annual electricity consumption, with the exception that changing EV consumption is associated with higher consumption levels (Figure AF.27). This was arguably to be expected, given that EVs are a significant source of extra electricity consumption for homes, so we would expect EV owners to be over-represented in higher-consumption homes.

Domestic Absence During Sessions. We also asked respondents "What best describes how you participated?" and provided them with two, non-mutually-exclusive options — i.e., was the respondent "already out of the house during the Session" and did the respondent "leave the house during the Session" — to which customers could respond "neither", "one", or "both". In posing these questions, we were especially curious as to whether respondents tended to opt into Sessions when they were already going to be out of the house. Such behavior may have counted as "flexibility" under NGESO's P376 preferred baselining methodology but it does not represent true reduction of energy consumption. Nevertheless, we found little evidence of this practice amongst survey respondents as just 9% said they were "already out of the house during the Session". This prevalence of domestic absence does not differ by remuneration level.

Spillover of Energy Behavior Into Periods "Around" Sessions. We also asked customers: (a) "Did participating in a Saving Session change what you did before the Session?"; and (b) "Did participating in a Saving Session change what you did after the Session?", obtaining the following responses:

- 60% of respondents said the Saving Session did not change their behavior before the Saving Session, whereas 23% of respondents and 13% of respondents respectively said that they used more electricity and used less electricity before the Saving Session.
- 50% of respondents said the Saving Session did not change their behavior afterwards, whereas 23% of respondents and 22% of respondents respectively reported using more electricity and less electricity after the Saving Session ended.

That the percentage of respondents who reported using less electricity was nine points higher after the Saving Session (i.e., 13% before the Session vs. 22% after) suggests that demand reduction behaviors may have continued past the Saving Session. Indeed, recall the small but nonzero effects of signing up to Saving Sessions on the half-hours "just after" and "just before" Sessions we see in our difference-in-differences regressions.

Note that that these percentages did not differ greatly in relation average remuneration per Session (Figures AF.23 and AF.24). However, higher-remunerated customers tended to report "no change" in consumption before or after the Saving Session, which may suggest that these customers were especially good at "concentrating" their consumption reduction during period wherein reduction is remunerated.

AI.3 Calculating the Wald estimator of impact of opt-in by hand versus through two-stage-least-squares regression DiD models

As discussed in Section 2.4.2, we obtained LATEs for our difference-in-differences designs using 2SLS regression and binary instruments in the form of multiplicative interactions between treatment group and treatment period. This is perhaps an unusual analytical strategy. Accordingly, we checked that that results obtained directly with our regression modelling reflects the basic logic of a complier average causal effect (CACE) as an ITT effect that is "diluted" by non-compliance. We did this by simply comparing the LATEs reported in the main text to Wald estimates calculated "by hand" by dividing our common and Session-specific ITT effects by compliance rates around Session participation. As shown in Tables AI.1 to AI.3, estimates obtained with the two methods are extremely close, where nonzero differences are likely due to the inclusion of average heating degree days as a covariate in our 2SLS regression models.

Table AI.1: Comparing LATE on opt-in from the Signed Up Early versus Late DiD to the Wald estimator calculated "by hand".

Model	ITT (kWh)	Opt-in rate	Wald estimator (WE) derived by hand (kWh)	LATE on opt-in (kWh)	Difference between WE by hand versus from LATE (kWh)
Overall common effect	-0.0897	63.0%	-0.1424	-0.1425	0.0001
November 15, 2022	-0.1172	70.5%	-0.1662	-0.1662	0.0000
November 22, 2022	-0.1194	64.3%	-0.1858	-0.1854	-0.0004
November 30, 2022	-0.1027	63.2%	-0.1624	-0.1624	0.0000
December 1, 2022	-0.0906	61.5%	-0.1473	-0.1472	-0.0001
December 12, 2022	-0.1009	67.4%	-0.1497	-0.1492	-0.0005
January 19, 2023	-0.0501	57.9%	-0.0865	-0.0860	-0.0005
January 23, 2023	-0.1337	72.2%	-0.1851	-0.1836	-0.0015
January 24, 2023	-0.1063	69.1%	-0.1538	-0.1527	-0.0011
January 30, 2023	-0.0515	63.9%	-0.0806	-0.0800	-0.0006
February 13, 2023	-0.0755	59.3%	-0.1273	-0.1271	-0.0002
February 21, 2023	-0.0754	65.1%	-0.1158	-0.1155	-0.0003
March 15, 2023	-0.0554	44.9%	-0.1233	-0.1223	-0.0010
March 23, 2023	-0.0503	59.4%	-0.0847	-0.0839	-0.0008

Note: We show a series of regression outputs. As labeled in the first column, the first model is the overall common effect, from our main regression model where the post-treatment period is all 29 half-hours when a Saving Session happened; the next 13 rows are from the regression models where the post-treatment period comprises only the half-hours from a specific Saving Session (see Section 2.4.2). In the second column, we show the ITT estimate from our Signed Up Early versus Late DiD, in kWh (Table AT.5 for the common effect, Table AT.2 for the individual Sessions). In the third column, we show the opt-in rate for the sample in that Session (or the average opt-in rate across all 13 Sessions, in the first model estimating the overall common effect). (These opt-in rates are different from Table 1 because the sample here is restricted to customers included in the regression; exclusions detailed in **??**.) In the fourth column, we show the Wald estimator derived from the division of the ITT by the opt-in rate, which is the equivalent of the compliance rate. In the fifth column, we show the LATE on opt-in derived from our 2SLS regression instrumenting the interaction between the indicator for Signed Up Early and the indicator for post-treatment (Table AT.6 for the common effect, Table AT.3 for the individual Sessions). In the sixth column, we show the difference in kWh between these two figures in the fourth and fifth columns. The fact that the differences are nonzero is due to the inclusion of average HDDs as a covariate in our DiD regressions; however, overall, the two methods of obtaining the impact of opt-in are in very close agreement.

Table AI.2: Comparing LATE on sign-up from the Octopus versus Bulb DiD to the Wald estimator calculated "by hand".

Model	ITT	Sign-up rate (Octopus)	Sign-up rate (Bulb)	Wald estimator (WE) derived by hand (kWh)	LATE on sign-up	Difference between WE by hand and LATE (kWh)
Overall common effect	-0.0361	36.7%	1.3%	-0.1021	-0.1021	0.0000
November 15, 2022	-0.0336	29.5%	0.0%	-0.1139	-0.1137	-0.0002
November 22, 2022	-0.0437	30.5%	0.0%	-0.1432	-0.1424	-0.0008
November 30, 2022	-0.0395	31.5%	0.0%	-0.1254	-0.1250	-0.0004
December 1, 2022	-0.0328	31.7%	0.0%	-0.1035	-0.1033	-0.0002
December 12, 2022	-0.0484	32.6%	0.0%	-0.1484	-0.1481	-0.0003
January 19, 2023	-0.0231	34.5%	0.0%	-0.0669	-0.0664	-0.0005
January 23, 2023	-0.0616	39.8%	0.0%	-0.1548	-0.1536	-0.0012
January 24, 2023	-0.0477	40.4%	0.0%	-0.1182	-0.1172	-0.0010
January 30, 2023	-0.0224	41.0%	0.0%	-0.0547	-0.0541	-0.0006
February 13, 2023	-0.0342	41.4%	0.1%	-0.0827	-0.0838	0.0011
February 21, 2023	-0.0306	41.5%	3.0%	-0.0796	-0.0812	0.0016
March 15, 2023	-0.0291	41.0%	6.3%	-0.0840	-0.0830	-0.0010
March 23, 2023	-0.0193	41.1%	7.3%	-0.0571	-0.0566	-0.0005

Note: We show a series of regression outputs. As labeled in the first column, the first model is the overall common effect, from our main regression model where the post-treatment period is all 29 half-hours when a Saving Session happened; the next 13 rows are from the regression models where the post-treatment period comprises only the half-hours from a specific Saving Session (see Section 2.4.2). In the second column, we show the ITT estimate from our Octopus v DiD, in kWh (Table AT.4 for the common effect, Table AT.1 for the individual Sessions). In the third column, we show the sign-up rate for Octopus customers in that Session (or the average sign-up rate across all 13 Sessions, in the first model estimating the overall common effect). In the fourth column, we show the sign-up rate for Bulb customers in that Session (or the average sign-up rate across all 13 Sessions, in the first model estimating the overall common effect). In the fifth column, we show the Wald estimator derived from the division of the ITT by the compliance rate, which in this case is the Octopus sign-up minus the Bulb sign-up rate. In the sixth column, we show the LATE on sign-up derived from our 2SLS regression instrumenting the interaction between the indicator for being an Octopus customer and the indicator for post-treatment (Table AT.4 for the common effect, Table AT.6 for the individual Sessions). In the seventh column, we show the difference in kWh between these two figures in the fifth and sixth columns. The fact that the differences are nonzero is due to the inclusion of average HDDs as a covariate in our DiD regressions; however, overall, the two methods of obtaining the impact of sign-up are in very close agreement.

Table AI.3: Comparing LATE on opt-in from the Octopus versus Bulb DiD to the Wald estimator calculated "by hand".

Model	ITT	Opt-in rate (Octopus)	Opt-in rate (Bulb)	Wald estimator (WE) derived by hand (kWh)	LATE on opt-in (kWh)	Difference between WE by hand versus from LATE (kWh)
Overall common effect	-0.0361	23.3%	1.0%	-0.1618	-0.1669	0.0051
November 15, 2022	-0.0336	20.8%	0.0%	-0.1616	-0.1613	-0.0003
November 22, 2022	-0.0437	19.5%	0.0%	-0.2239	-0.2222	-0.0017
November 30, 2022	-0.0395	19.8%	0.0%	-0.1999	-0.1992	-0.0007
December 1, 2022	-0.0328	19.4%	0.0%	-0.1694	-0.1691	-0.0003
December 12, 2022	-0.0484	21.8%	0.0%	-0.2226	-0.2221	-0.0005
January 19, 2023	-0.0231	19.6%	0.0%	-0.1182	-0.1173	-0.0009
January 23, 2023	-0.0616	28.1%	0.0%	-0.2192	-0.2174	-0.0018
January 24, 2023	-0.0477	27.4%	0.0%	-0.1741	-0.1725	-0.0016
January 30, 2023	-0.0224	25.2%	0.0%	-0.0891	-0.0882	-0.0009
February 13, 2023	-0.0342	23.2%	0.0%	-0.1478	-0.1506	0.0028
February 21, 2023	-0.0306	25.9%	2.2%	-0.1291	-0.1327	0.0036
March 15, 2023	-0.0291	17.6%	3.3%	-0.2026	-0.1999	-0.0027
March 23, 2023	-0.0193	23.6%	4.8%	-0.1023	-0.1015	-0.0008

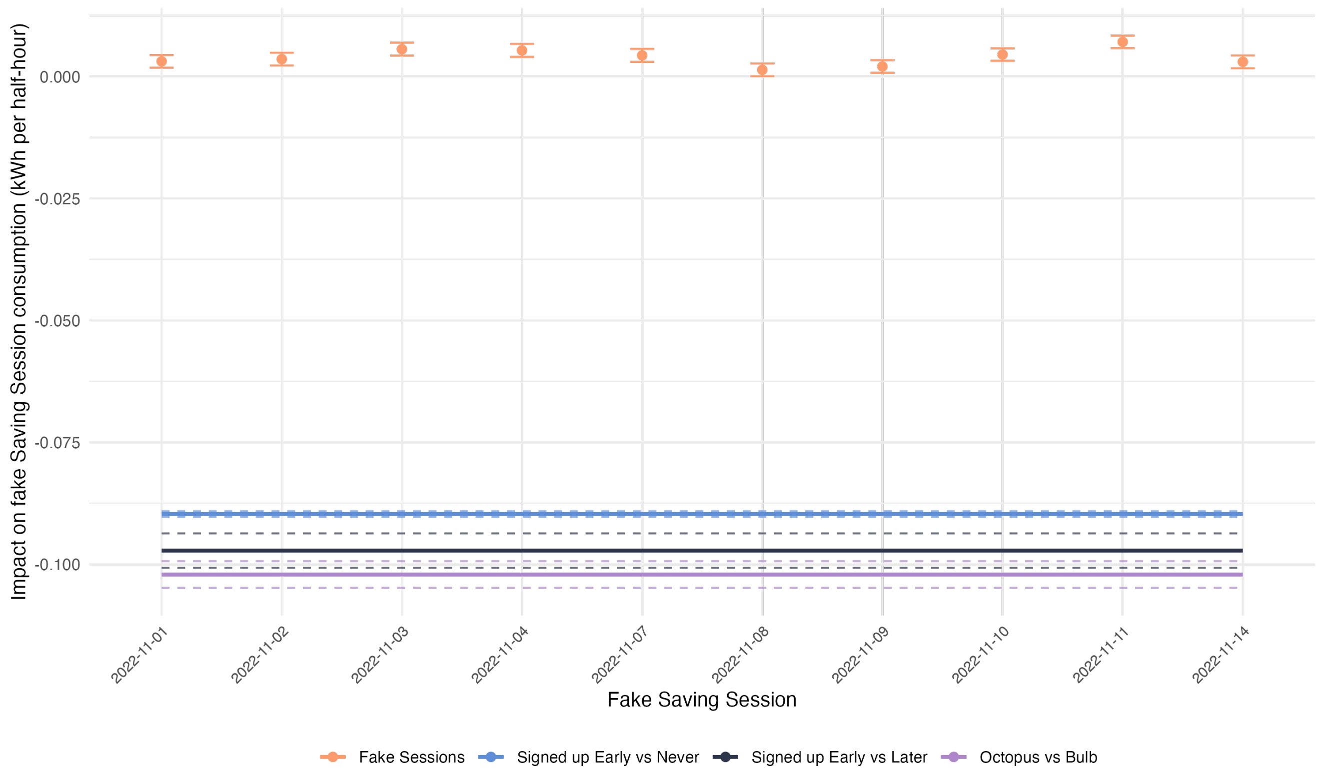
Note: We show a series of regression outputs. As labeled in the first column, the first model is the overall common effect, from our main regression model where the post-treatment period is all 29 half-hours when a Saving Session happened; the next 13 rows are from the regression models where the post-treatment period comprises only the half-hours from a specific Saving Session (see Section 2.4.2). In the second column, we show the ITT estimate from our Octopus versus DiD, in kWh (Table AT.5 for the common effect, Table AT.3 for the individual Sessions). In the third and fourth columns, we show the opt-in rates for Octopus and Bulb customers in that Session (or the average opt-in rate across all 13 Sessions, in the first model estimating the overall common effect). (These opt-in rates are different from Table 1 because the sample here is restricted to customers included in the regression; exclusions detailed in Figure AF.14.) In the fifth column, we show the Wald estimator derived from the division of the ITT by the compliance rate, which in this case is the Octopus opt-in minus the Bulb opt-in rate. In the sixth column, we show the LATE on opt-in derived from our 2SLS regression instrumenting the interaction between the indicator for being an Octopus customer and the indicator for post-treatment. In the seventh column, we show the difference in kWh between these two figures in the fifth and sixth columns. The fact that the differences are nonzero is due to the inclusion of average HDDs as a covariate in our DiD regressions; however, overall, the two methods of obtaining the impact of opt-in are in very close agreement.

AI.4 DiD Placebo Tests

Placebo tests to check parallel trends. We conducted a series of placebo tests to formally test parallel trends. For our main DiDs, our "pre" period comprises the weekdays during October 2022 and the first 14 days of November 2022. In our placebo tests, we restrict our pre-treatment period to weekdays during October 2022 only. We define a series of fake Saving Sessions, each of which comprises the "post" period in each placebo test. We define the following fake Sessions, each of which we code to have occurred 17:00 to 18:00 (1 hour), across each weekday in November before the first real Saving Session on November 15, 2022 (November 1, 2, 3, 4, 7, 8, 9, 10, and 11)

We use the Signed Up Early versus Never DiD groups for these placebo tests. In Figure AI.1, we show the point estimates and 95% confidence intervals for each of the nine placebo DiDs (in orange). All of the fake Sessions have difference-in-differences above zero (and confidence intervals that do not cross zero). However, the effect sizes of 0.0071 to 0.0014 are much smaller than the effect sizes we see in our DiDs of real Sessions (see Table AT.4) of -0.0897, -0.0972, and -0.1021 from signed-up customers per half-hour. In Figure AI.1, we show the effects of sign-up estimated from each of our three DiDs (see Table AT.4) for scale. We believe the small positive DiD is evidence of slight differences in trends between the Signed Up Early versus the Never Signed Up groups (in comparing November 17:00 to 18:00 consumption to October 09:00 to 22:00 consumption.)

Figure AI.1: Coefficient plot for nine placebo tests.

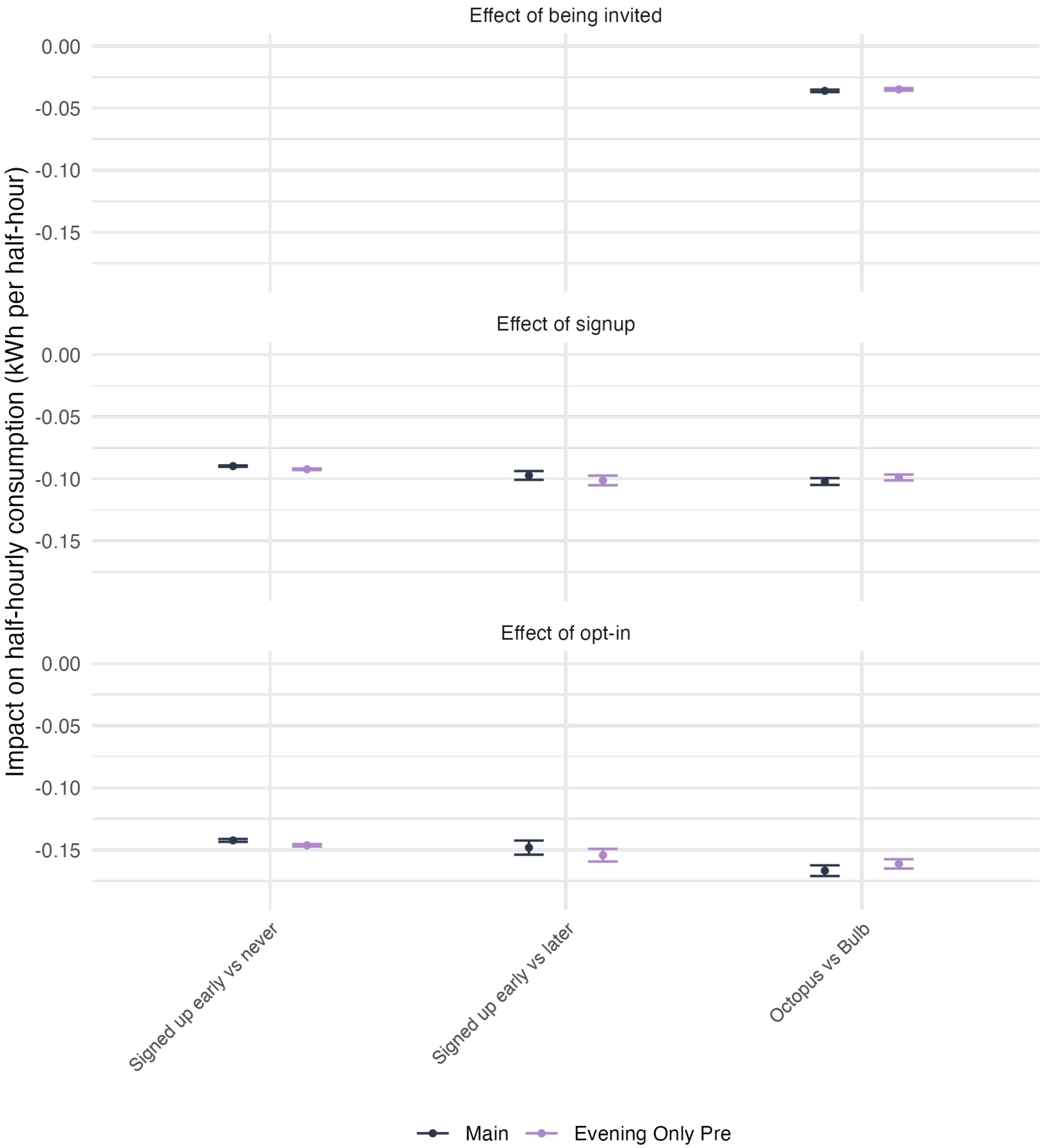


Note: Coefficient (and 95% confidence intervals) on the difference-in-differences coefficient in a series of DiDs for each of nine fake Saving Session regressions in November, where the post-treatment period in each regression is customers' half-hourly consumption during each of the nine fake Saving Sessions. We use the Signed Up Early versus Never DiD groups for these placebo tests. We show the effects of sign-up estimated from each of our three DiDs (see Table AT.4) for scale; their solid lines are the point estimates from each DiD, and the dotted lines each DiD's 95% confidence interval upper and lower limits.

AI.5 DID alternative pre-treatment period specification

Examining how DiD results change when using a different pre-treatment period. The pre-treatment period in our DiDs comprised all half-hours between 09:00 to 22:00 from October 1, 2022 through November 14, 2022. We re-ran the DiDs using a pre-treatment period comprised of half-hours between *16:00 to 19:00* on the same days (October 1, 2022 through November 14, 2022). We show both sets of DiDs' results in Figure AI.2. The results are similar between the two specifications of the pre-treatment period.

Figure AI.2: Coefficient plot for nine pre-trends placebo tests.



Note: In the first two DiDs (Signed Up Early versus Never, and Signed Up Early versus Late), the effect of sign-up is an ITT estimate, while the effect of opt-in is a LATE. Octopus versus Bulb's ITT estimate is the effect of being invited; the effect of sign-up and of opt-in are LATEs. The main specification (in navy) pre-treatment period is weekday half-hours from 09:00 to 22:00 in October 2022 and the first two weeks of November 2022. The alternate specification (in purple) pre-treatment period is weekday half-hours from 16:30 to 19:30.

AI.6 Examining how our DiD results compare to the official NGESO "pre-post" methodology (i.e., "P376").

In Tables AI.4 and AI.5, we show how our DiD results compare to the official methodology endorsed by NGESO (i.e., "P376"). As described in Section 2.4.1, there are two versions of the P376 methodology, "unclipped" (a modified pre-post methodology) and "clipped" (where demand *increases* are clipped to 0).

Table AI.4: Demand reduction (kWh per half-hour) according to our DiDs, unclipped P376, and clipped P376.

Group	N	Demand reduction from DiD	Avg unclipped P376 demand reduction (% difference from DiD)	Avg clipped P376 demand reduction (% difference from DiD)
Signed up never	654062		-0.0008	0.1132
Signed up early	332195	0.0897	0.0927 (3.4%)	0.1664 (85.6%)
Signed up late	12438		-0.0030	0.1138
Signed up early	331992	0.0972	0.0995 (2.4%)	0.1696 (74.5%)
Bulb	197307		0.0010	0.1186
Octopus	1137028	0.0361	0.0312 (-13.7%)	0.1314 (263.9%)

Note: We show the difference-in-difference coefficient from our three DiDs' ITT analyses, compared to the same analyses' "control" and "treatment" groups' demand reduction as measured by the unclipped and clipped versions of NGESO's prescribed methodology for DFS providers, shorthanded as "P376". In parentheses after the NGESO methodology estimates, we show the percent difference compared to our DiDs' estimates. We see minimal bias from the unclipped version of this method. When examining the full treatment groups in our DiDs, we see substantial bias from the clipped P376 method. However, this bias mostly comes from customers who have not opted in to Sessions in our DiDs' treatment groups.

Table AI.5: Demand reduction (kWh per half-hour) among opt-ins only according to DiDs, unclipped P376, and clipped P376.

Group	N	Demand reduction from DiD (LATE on opt-in)	Avg unclipped P376 demand reduction (opt-ins only)	Avg clipped P376 demand reduction (opt-ins only)
Signed Up Early versus Never	332195	0.1425	0.1311 (-8.0%)	0.1843 (29.3%)
Signed Up Early versus Late	331992	0.1483	0.1381 (-6.9%)	0.1857 (25.2%)
Octopus versus Bulb	1137028	0.1669	0.1273 (-23.7%)	0.1817 (8.9%)

Note: We show the difference-in-difference coefficient from our three DiDs' local average treatment effect on *opt-in*, compared to the average demand reduction as estimated by the NGESO "P376" methodology for opted in customers from each DiD's "treatment" group. We show two versions of the NGESO methodology: unclipped and clipped. In parentheses after the NGESO methodology estimates, we show the percent difference compared to our DiDs' estimates. Our DiDS' LATE on opt-in are on average higher than demand reduction estimated by *unclipped* P376 among opt-ins, though still lower than the demand reduction estimated by *clipped* P376 among opt-ins.

AI.7 Regression Discontinuity Design: Extended Details

AI.7.1 Background: Known Assignment Mechanism, But No Randomization

Like the other Demand Flexibility Service (DFS) events delivered by Octopus Energy throughout the Winter of 2022-23, Octopus Energy customers who agreed to participate in the Saving Session on February 13 should have received a notice the day prior. However, notices were delayed (Figure 8). And they were ultimately sent in accordance with the ordering of customers' account IDs (i.e., a string of integers ranging in length) which are, in turn, a function of each customer's *tenure* — where customers new to Octopus Energy generally have account IDs that are larger in magnitude (Figure AF.16).

To clarify, opt-in notices for Saving Sessions were distributed to customers in scheduled batches using a roster of account IDs. Batching is a standard practice used to minimize error in the delivery of messages to a large number of customers. And, as a general rule, Octopus Energy does not release batched communications to customers between 20:00 and 8:00. Nevertheless, the process by which messages to customers are generated, batched, and sent is inexact. This is owing to idiosyncratic server delays — where opt-in notices for the Saving Session on February 13 were deferred to an unusual degree into the evening of the 12th and through to the morning of the 13th (Figure 8). Thus, some DFS-participating Octopus Energy customers received day-ahead opt-in notices whereas others received an opt-in notice for the February 13 Saving Session sometime before 1PM on the day of the Saving Session itself (i.e., intraday notice).

Note well that account IDs were not manually batched. Instead, a batch size was first manually chosen. And then the internal platform Octopus Energy uses for external communication was scheduled to dispense batches of notices in order of the magnitude of customers' account IDs. Still, owing to the batching process and standard server lag, account ID does not strictly (i.e., monotonically) increase with time (Figure AF.16).[57] Furthermore, owing to the above-mentioned imperfect nature of message delivery in relation to Octopus Energy's customer-communication platform, batched messages began to be sent again around 7:45 on February 13 (Figure 8). Nevertheless, we use 8:00 as our temporal cutoff as this is the point at which Octopus Energy formally allows customer contact.

AI.7.2 Specification and Bandwidth Choice Given 'Donut-Hole' RDD

To fit the models for our regression discontinuity design (RDD), we only use a 1st-degree polynomial (i.e., a linear fit) for our centered assignment variable $(A_i - C)$ and a multiplicative interaction between $(A_i - C)$ and our binary treatment to allow "local" regression lines on either side of our cutoff for treatment C. We prefer this simple model specification in light of warnings of overfitting and nonsensical conclusions when performing regression discontinuity using higher-order polynomials (see Huntington-Klein (2021, p. 516-518) as well as Gelman and Imbens (2019), Gelman and Zelizer (2015)).

Indeed, we prefer a *less*-flexible fit reflective of information further from the cutoff C. This owing to our use of "donut-hole" RDD (Barreca et al., 2011, 2016) whereby we must necessarily *extrapolate* the regression line forward across the region of our account-ID-based running variable to the immediate left of our cutoff for which we exclude all Octopus Energy customers sent overnight notices (Figure 8). For this reason, we also eschew weighting observations with values for the assignment variable closer to the threshold during model fitting. Instead, all observations are weighted equally (i.e., a "uniform kernel") and we simply present models using multiple bandwidths (wide and narrow) following the recommendations of Lee and Lemieux (2010).

More formally, given our use of donut-hole RDD, we estimate one *pair* of optimal *asymmetric* bandwidths h using the techniques of Cattaneo and colleagues (Calonico et al., 2017, Cattaneo et al., 2019, Forthcoming, Cattaneo and Vazquez-Bare, 2017) as implemented in the newer, Python-based version of their popular STATA function "rdbwselect" (Calonico et al., 2017). The pair of bandwidths is referred to as "optimal" as it is automatically estimated given: (a) the data; (b) a polynomial order for points estimation and bias correction (here, respectively, the defaults of 1st- and 2nd-degree); (c) a kernel weighting function (here, uniform), and (d) a means of calculat-

[57] Details of the message-delivery process obtained during conversation between the authors of this research and technical experts at Octopus Energy Group.

ing variance (here, K-nearest-neighbours, where $K = 3$, the default value), amongst other factors (Cattaneo and Vazquez-Bare, 2017). We limit our attention to a pair of optimal bandwidth expected to minimize mean-squared error or "MSE" (i.e., the average of the squared deviations between predicted and observed values). However, we probe the sensitivity of our results by fitting ancillary models wherein we expand each MSE-optimal bandwidth by a factor of 1.5 and a factor of 2 (both arbitrarily chosen). We use "rdbwselect" to obtain our pair of MSE-Optimal bandwidths without the use of pre-treatment covariates.

Note that an optimal bandwidth is specific to an outcome variable (Cattaneo and Vazquez-Bare, 2017, p. 143). Thus, we use two pairs of asymmetric MSE-optimal bandwidths. The first pair is specific to our models of in-Session consumption ($h_{Left,\text{ MSE-Optimal, Consumption}}$ and $h_{Right,\text{ MSE-Optimal, Consumption}}$), and the second is specific to our models of Session participation ($h_{Left,\text{ MSE-Optimal, Opt-in}}$ and $h_{Right,\text{ MSE-Optimal, Opt-in}}$).

To actually filter our data, we construct a range of valid account IDs by taking our constructed ID-based threshold $C = 2,454,839$ and subtracting h_{Left} and adding h_{Right}. The bandwidths obtained using "rdbwselect" are as follows:

- $h_{Left,\text{ MSE-Optimal, Consumption}} = 244,339.92$
- $h_{Right,\text{ MSE-Optimal, Consumption}} = 495,274.84$
- $h_{Left,\text{ MSE-Optimal, Opt-in}} = 489,367.98$
- $h_{Right,\text{ MSE-Optimal, Opt-in}} = 448,440.77$

And, to show how our results behave under different ranges of our running variable, we expand these bandwidths as follows:

- $h_{Left,\text{ MSE-Optimal, Consumption}} = (244,339.92 \div 1.5)$
- $h_{Right,\text{ MSE-Optimal, Consumption}} = (495,274.84 \times 1.5)$
- $h_{Left,\text{ MSE-Optimal, Consumption}} = (244,339.92 \div 2)$
- $h_{Right,\text{ MSE-Optimal, Consumption}} = (495,274.84 \times 2)$
- $h_{Left,\text{ MSE-Optimal, Opt-in}} = (489,367.98 \div 1.5)$
- $h_{Right,\text{ MSE-Optimal, Opt-in}} = (448,440.77 \times 1.5)$
- $h_{Left,\text{ MSE-Optimal, Opt-in}} = (489,367.98 \div 2)$
- $h_{Right,\text{ MSE-Optimal, Opt-in}} = (448,440.77 \times 2)$

Finally, note that we observe balance on relevant pre-treatment covariates (e.g., historical energy usage) when using our MSE-optimal bandwidth (see Figure AI.3).

AI.8 Randomized Field Trial: Extended Details

AI.8.1 Background: Random Assignment Mechanism, But Non-Compliance

Random assignment to our first experimental condition (i.e., day-ahead email) was uncomplicated. However, random assignment to our second condition (i.e., SMS plus bonus) suffers from non-compliance. Accordingly, threats to our ability to credibly estimate the causal effect of the SMS-plus-bonus condition include: (a) unobserved third factors determinant of both energy consumption and an Octopus Energy customer's decision to allow SMS communications from Octopus (e.g., socio-economic status; environmental conscientiousness); and (b) residual imbalance of potential outcomes stemming from our random sub-sampling (Gelman et al., 2020).[58]

[58] We observe good balance on relevant pre-treatment covariates across the three experimental conditions (Figure AI.4).

Recall from the main text that, given non-compliance, we adopt the framing of a randomized encouragement design (RED) — i.e., a type of experimental setup wherein variation in some difficult-to-directly-manipulate treatment is induced using a source of random variation (i.e., the random ”encouragement“) that is related to the difficult-to-directly-manipulate treatment *and not* related to the outcome of interest. For example, consider randomized assignment of advertisements across geographic markets and hours of watching a video-streaming service platform, like Netflix or Hulu, in an analysis relating the probability of service renewal to hours of television watched, the latter of which may be difficult to control.

As discussed by Gelman et al. (2020), REDs allow one to estimate causal effects for participants whose behavior *could be altered* by the random encouragement. These individuals are known as ”compliers“ — i.e., a latent subpopulation of entities and the only individuals for whom a RED provides counterfactual information. This is because the treatment-related behavior of compliers could differ under heterogeneous experimental conditions. Using the above streaming example for instance, compliers would be the group of people for whom differential exposure to advertisement could results in different amounts of television watched.

Along this line, REDs allow one to recover a special kind of local average treatment effect (LATE) known as the complier average causal effect (CACE). Importantly, the CACE is distinct from and, possibly, unrepresentative of the average treatment effect (ATE) in both the broader sample and the wider population from which that sample is drawn (Basu et al., 2018, Gelman et al., 2020). For our field experiment and our treatment with non-compliance, the ATE is the overall effect of receiving an intraday SMS reminder and being made bonus-incentive eligible.

AI.8.2 Consideration of Instrumental Variables (IV) Assumptions

IV estimation is sometimes viewed with scepticism due to its multiple, peculiar assumptions that are distinct from the familiar (conditional) ignorability assumption. For example, see Gelman et al. (2020, ch. 21), Hernán and Robins (2023, ch. 16), and Cunningham (2021, p. 315-386). Accordingly, here we consider the formal assumptions of IVs in relation to our field trial.

Recall that all individuals assigned to receive our second treatment were made eligible for the bonus price incentive such that the only dimension along which compliance could vary is receipt of the SMS head-up. Accordingly, for our field trial there can be no *always takers* (i.e., Octopus Energy customers who would have received an intraday SMS reminder even if they had not been randomly assigned to the SMS-plus-bonus condition). This is by design as we randomly selected which Octopus Energy customers were sent SMS text messages. Customers were fully unaware of this randomization, only learning of the SMS reminder at the time that they received it — if they ever received it. And it was impossible for customers not assigned to the SMS-plus-bonus condition to be inadvertently sent an intraday SMS reminder. Thus, our design also precludes *defiers* — i.e., individuals whose treatment status is always opposite of their randomized encouragement (e.g., customers who would manoeuvre to receive an intraday SMS when not assigned to the SMS group and customers who would manoeuvre to never receive a SMS when assigned to the SMS group). Nevertheless, our design does not rule out *never takers* as there are customers who would not receive an intraday SMS no matter their random assignment (e.g., those customers who disallow SMS communications from Octopus Energy; customers who inadvertently agree to receive Octopus Energy's SMS communications but who screen for ”spam“ to automatically block texts from unknown senders/phone numbers). As discussed by Gelman et al. (2020, p. 425), the presence of never-takers exacerbates violations of the exclusion restriction such that the CACE is expected to be biased by α (i.e., the associated between the instrument Z and the outcome y for the never-takers) multiplied by a factor stemming from the ratio of the proportion of never-takers and the proportion of compliers. Here, this ratio is expected to be at least 1.05 — i.e., 4,731 $\div$ 4,472 — where this first figure is the number of individuals randomly assigned to the SMS-plus-bonus condition but who had disallowed SMS messages from OE.

Beyond the assumptions related to the aforementioned sub-populations, there are five primary assumptions of IV estimation (Gelman et al., 2020, p. 421-426, Section 21.1) in addition to a sixth assumption (Bhuller and Sigstad, 2022) resulting from our decision to estimate the effect of two binary treatments in the same model.

1. Ignorability of the Instrument With Respect to Treatment Potential Outcomes (T^0, T^1) and With Respect to the Response Potential Outcomes (y^0, y^1): Recall that potential outcomes are one's value for a response variable (e.g., session consumption) when exposed to different experimental conditions at the same point in time. And this assumption implies that our instrument (i.e., assignment to the SMS-plus-bonus condition) is independent of the potential outcomes for actually receiving the SMS-plus-bonus treatment *and* the potential outcomes for our response variables (i.e., kWh of energy consumption; participation in the March 15 Savings Session). This is clearly satisfied as our assignment to the SMS-plus-bonus condition was randomized. Thus we do not anticipate third factors determinant of condition assignment and actual receipt of the SMS treatment or third factors determinant of condition assignment and either of our two response variables (i.e., session consumption, session participation).

2. Monotonicity: This assumptions implies that there are no *defiers*. As discussed above, this is precluded by the design of our study as Octopus Energy customers were unaware of our randomization, Octopus Energy customers in the control condition (i.e., intraday notice only) could not themselves join the SMS-plus bonus-condition, and Octopus Energy customers who actually received the SMS-plus-bonus treatment could not somehow reject it to join the control group.

3. Nonzero Association Between the Instrument and the Treatment: This assumption implies that the covariance of the instrument and the treatment is not equal to zero. This assumption is testable. And our models (Sections AT.7 and AT.8) unsurprisingly indicate a positive association between random assignment to the SMS-plus-bonus condition and receiving the SMS-plus-bonus treatment.

4. Exclusion Restriction: This assumption implies that the randomized instrument (i.e., assignment to the SMS-plus-bonus condition) is not directly associated with the outcome variable (i.e., session consumption and session participation) for never-takers and always-takers (Gelman et al., 2020, p. 423). This assumption is trivially satisfied in our case as it was impossible for Octopus Energy customers to receive the SMS-based-treatment on their own (i.e., always-takers) or to independently learn about and ultimately reject the SMS-based treatment (i.e., never-takers).

5. No Crossed Effects (Bhuller and Sigstad, 2022): This assumption implies that our randomized instrument is not associated with the other treatments in the model (here, the day-ahead-heads-up-email condition). This assumption is easily satisfied in our case as our two treatment groups are distinct by design and because the day-ahead-heads-up-email condition has (to our knowledge) perfect compliance.

Given all of this, we maintain that instrumental variables estimation can be defensibly used to recover the CACE of the SMS-plus-bonus condition on consumption during, and participation in, the March 15 Saving Session.

AI.9 Balance on Pre-Treatment Covariates (RDD and RCT)

As discussed by Gelman et al. (2020), under a strict interpretation, balance relate to similarity in the *distributions* of potential-outcomes-relevant pre-treatment variables across levels of a treatment variable — not merely summary statistics (e.g., the mean). Moreover, hypothesis tests for ”statistically significant“ differences across treatment groups in experiments and quasi-experiments have been subject to repeated critique under the view that balance is a property of a give sample, not the population from which it is drawn (Harvey, 2018, Imai et al., 2008, Senn, 1994).

Accordingly, we eschew comparisons like t-tests to instead qualitatively compare distributions of pre-treatment covariates across experimental groups using quantile-quantile (Q-Q) plots, following Imai et al. (2008). As our concern is energy-consumption related behavior, we narrowly focus on historical energy usage — i.e., Total P376 (Unadjusted) Baseline (kWh) and Estimated Annual Consumption (kWh). As mentioned in the main text, the former is an unweighted average of consumption during the same half-hour of the day during the ten most-recent working days as governed by the the P376 amendment to Great Britain's electricity balancing and settlement code. And the latter is Octopus Energy's predicted customer consumption based on meter readings over one year.

Given the large size of our data, which are drawn from across Great Britain, we create Q-Q plots specific to region. For this reason, we also consider balance on the degree to which an Octopus Energy customer's *postcode* is deprived using an index of multiple deprivation (IMD) which combines, in a weighted manner, multiple facets of poverty (e.g., crime, barriers to housing, health, amongst other factors).[59] Last, for our RDD, we also include Q-Q plots of account ID and customer tenure to show the by-design discontinuity of the former and the sneaking overlap of the latter.

[59] For further details, see the UK Ministry of Housing, Communities & Local Government.

Figure AI.3: Small-multiple Q-Q Plot for balance of relevant pre-treatment covariates across non-randomized experimental groups depending on MSE-optimal bandwidth and outcome (i.e., session consumption or session participation). For the Index of Multiple Deprivation (IMD), more deprived areas have *lower* postcode ranks.

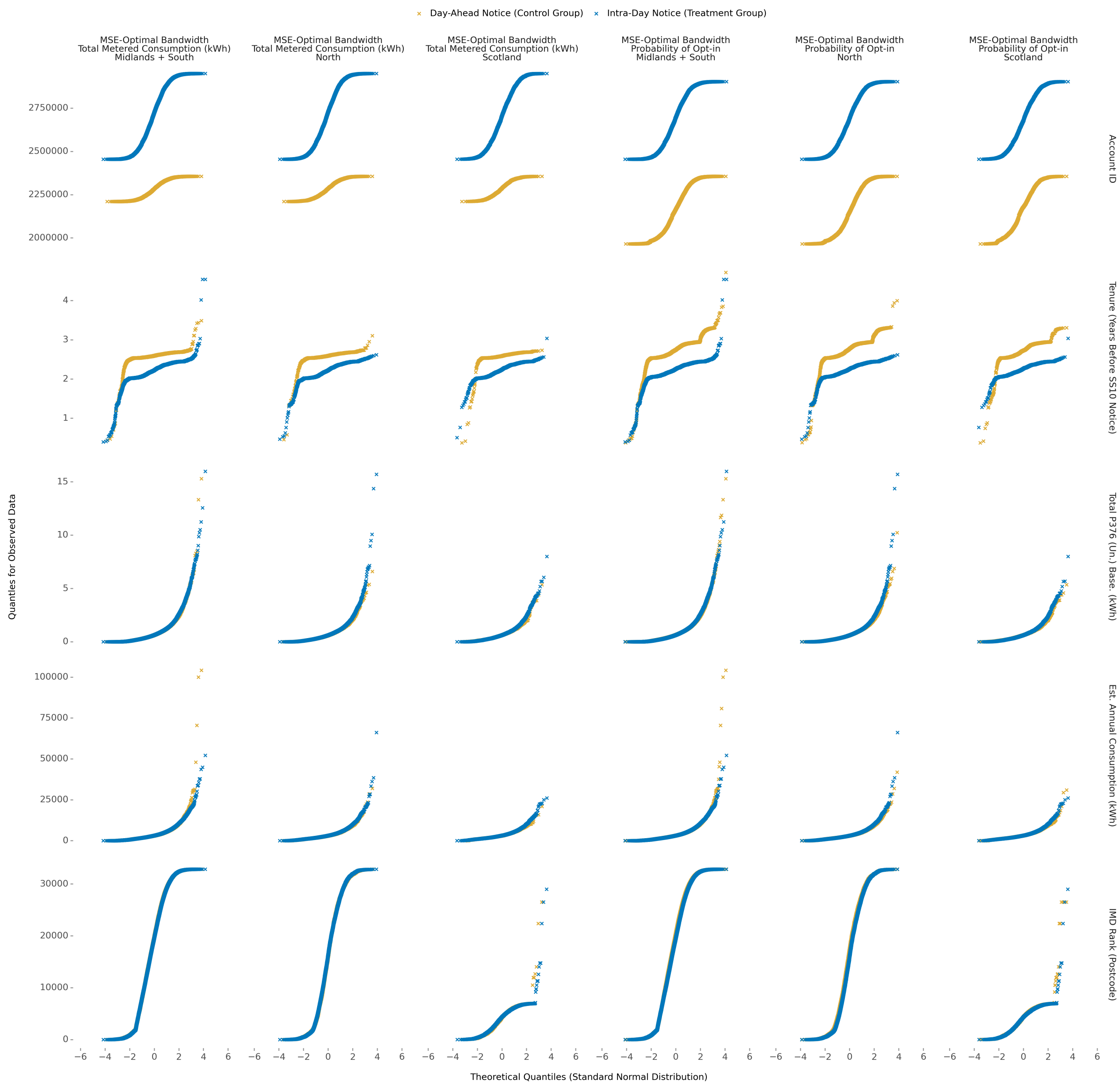
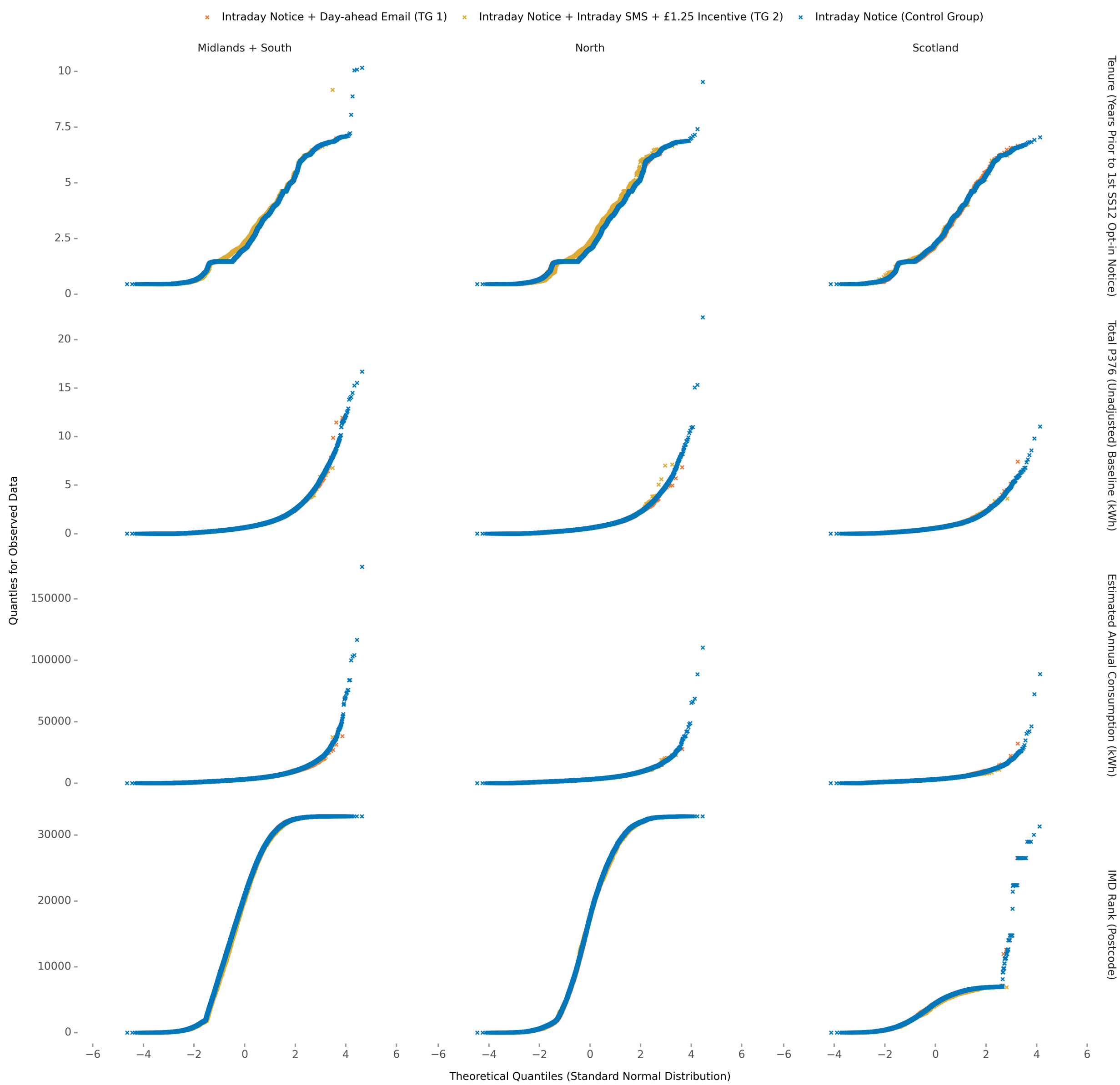


Figure AI.4: Small-multiple Q-Q Plot for balance of pre-treatment covariates across randomized experimental conditions. For the Index of Multiple Deprivation (IMD), more deprived areas have *lower* postcode ranks.



AI.10 Simple Linear Regression and Limited Response Variables (RDD and RCT)

Recall that we use ordinary-least-squares (OLS) regression models of our two response variables: (a) total electricity usage during a 60 minute Saving Session (kWh); and (b) participation in a given Saving Session (i.e., opt-in). The latter is of course binary, and consumption is a non-negative, real-valued quantity with true (i.e., valid) zeros (n.b., some customers' may use stored energy captured via their solar panels and retained with a domestic battery).

Accordingly, it may seem odd that we use simple linear regression to model two limited (i.e., range-restricted) variables as opposed to, for instance, Gamma regression for our continuous (skewed) positive outcome and logistic regression for our binary outcome. This choice may seem especially odd to readers unfamiliar with econometric techniques or (Bayesian) readers who maintain that a statistical model ought to encapsulate theoretical mechanisms such that it is *generative* of the observed data (e.g., see Gabry et al. (2019)).

Nevertheless, OLS regression is the standard technique for obtaining causal effects amongst econometricians. These scientists tend to narrowly focus on accurately estimating treatment effects in the form of conditional means as opposed to performing in-sample or out-of-sample prediction tasks. Comprehensive discussion of this OLS-centric logic is found in various text (Angrist, 2006, Angrist and Pischke, 2009, Basu et al., 2018, Wooldridge, 2010). We adopt this approach in this study.

That said, use of simple linear regression to model limited dependent variables is not without controversy — particularly with respect to instrumental variable (IV) estimation with a binary instrument, binary treatment, and a binary outcome (Section AI.8). Indeed, researchers from multiple scientific disciplines, including economics, have cautioned against using linear regression in this scenario (Bhattacharya et al., 2006, Dong and Lewbel, 2015, Hollenbach et al., 2019, Huntington-Klein, 2021, Kleibergen and Zivot, 2003, Li et al., 2022).

Still, we prefer OLS given its popularity (Angrist, 2006, Angrist and Pischke, 2009, Wooldridge, 2010) and easy interpretation, where linear probability models enjoy a number of interpretive benefits compared to *non*-linear probability models (Battey et al., 2019, Breen et al., 2018, Gomila, 2021, for comparisons, see). Moreover, we anticipate that our results are robust to use of simple linear regression following methdological discussions by economists, psychologists, and sociologists (Angrist, 2006, Basu et al., 2018, Breen et al., 2018, Wooldridge, 2010), particularly given the large number of observations used to fit our models.

Indeed, Chiburis et al. (2012) suggest that, in the binary-instrument-binary-treatment-binary-outcome scenario, results from IV estimation obtained using linear probability models and results from a probit model are only likely to differ for sample sizes less than 5,000 and when the probability of treatment and/or the probability of a positive outcome for the response (here, session opt-in) are close to 0.1 or 0.9. This latter warning is perhaps relevant for our RED as roughly 23% of individuals received the SMS-plus-bonus treatment. However, Basu et al. (2018) point to the favourable performance of two-stage least-squares for estimating the LATE in the binary-instrument-binary-treatment-binary-outcome scenario across a wide range of treatment rarities, where we note again that our sample sizes are substantial.

Given all of this, we opt for simple OLS. However, in so doing, we necessarily adopt an interpretation of our models as linear projections of our nonlinear outcomes (Gomila, 2021, Wooldridge, 2010), where we note the fundamental heteroscedasticity of linear probability models (Rencher and Schaalje, 2008, p. 508-509). Furthermore, we limit our attention to point estimates as opposed to predicted quantities across a wide range of values for our covariates which could, in some instances, be nonsensical (e.g., negative consumption or negative probabilities). Along this line, all continuous pre-treatment variables and possible confounders enter our regression equations as Z-scores such that coefficients in our expanded model reflect covariates held at their global average values as opposed to values at the extreme (see Wooldridge (2010, p. 563)).

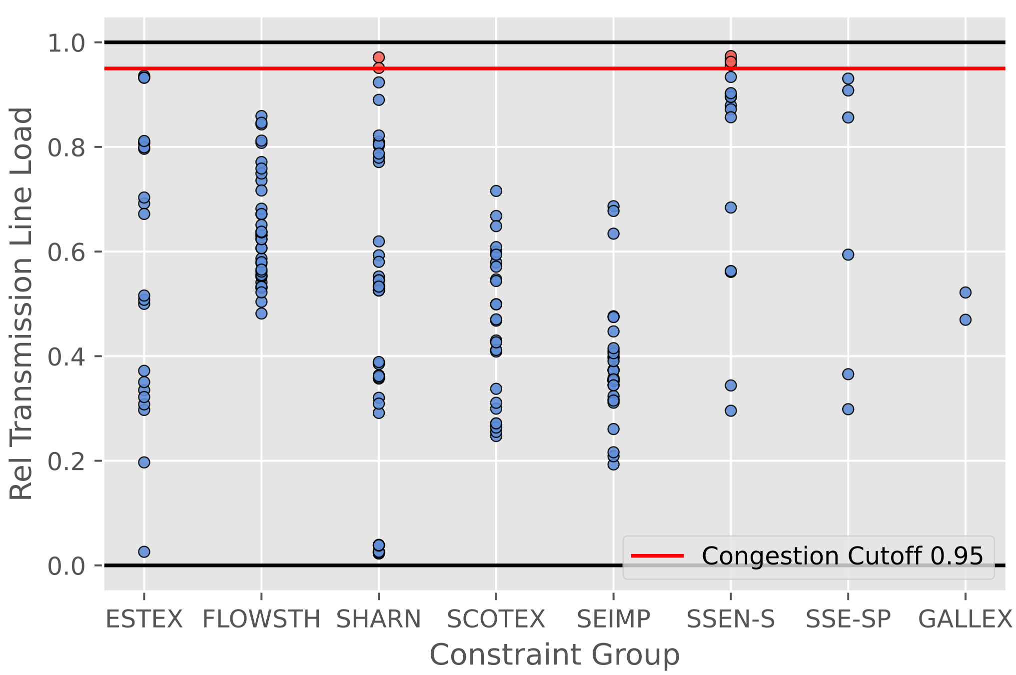
As discussed by Gelman et al. (2020, p. 154-155), the outcome variable is not required to be normally-distributed when using simple linear regression and both the unequal variance of errors (i.e., heteroscedasticity) and the non-normality of errors are expected to be minor issues when using simple linear regression to estimate conditional means and to recover the regression line as opposed to the prediction of individual data points. Still, to try to counter some of these issues, all of our models use heteroscedasticity-consistent standard errors, specifically the HC0 variant (see Long and Ervin (2000) for a comparison heteroscedasticity-consistent covariance matrices).

AI.11 Assumptions About Electricity Transmission Constraints During DFS Events

DFS events reduced the generation required to by NGESO to ensure sufficient reserve capacity. To estimate their cost-effectiveness against alternative options available to NGESO, as well as their welfare impacts, we first identify the most likely prevented fuel. At times when transmission lines are constrained, it is possible that demand reduction in areas of the country with sufficient reserve would be less useful than demand reduction in areas of the country without sufficient reserve. Under these conditions, Great Britain's electricity market should no longer be considered as having a national marginal generator, but instead multiple regional ones.

For example, grid congestion mostly occurs between Scotland and England, and is usually caused by excess wind generation (Savelli et al., 2022) in Scotland. Local electricity savings '*behind*' those boundaries, in Scotland, are unlikely to reduce carbon emissions to the same extent that savings in England would. With this in mind, we estimated to what extent Great Britain's transmission grid was operating at capacity during the 13 Saving Sessions. In Figure AI.5 we show day-ahead relative line loads of Great Britain's main constraint groups, each point representing the 29 half hours during Saving Session (data taken from National Grid (2023*a*)). We call a line *congested* when a relative line load exceeds a threshold of 0.95.

Figure AI.5: Power flow across main constraint groups in the GB transmission grid represented as flow relative to physical limit during Saving Sessions.



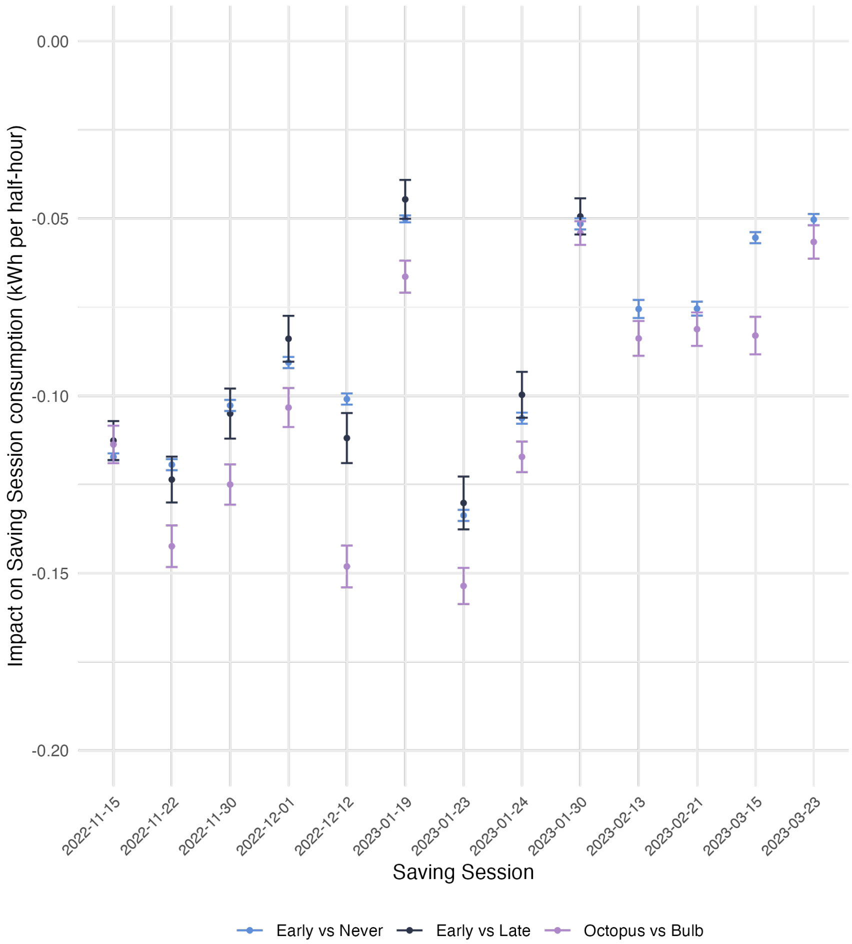
Note: Over the 29 half-hour periods of the 13 Saving Sessions, data from day-ahead relative line loads of Great Britain's main constraint groups shows that only 6 half hours were considered to be "congested" -assuming that the transmission lines are "congested" when the relative transmission line load is above 95% of its capacity-. If defined as above, only the SHARN and SSEN-S constrain groups show congestion during the Saving Sessions, while the other lines show relative transmission line load between 20 and 80% in most of the instances.

We found that during the 29 half-hours when Saving Sessions occurred, only 6 half-hours were congested. Given this low rate of congestion, and for simplicity, in our cost-effectiveness and welfare analyses, we considered Great Britain as having a national marginal generator.

AF Appendix Figures

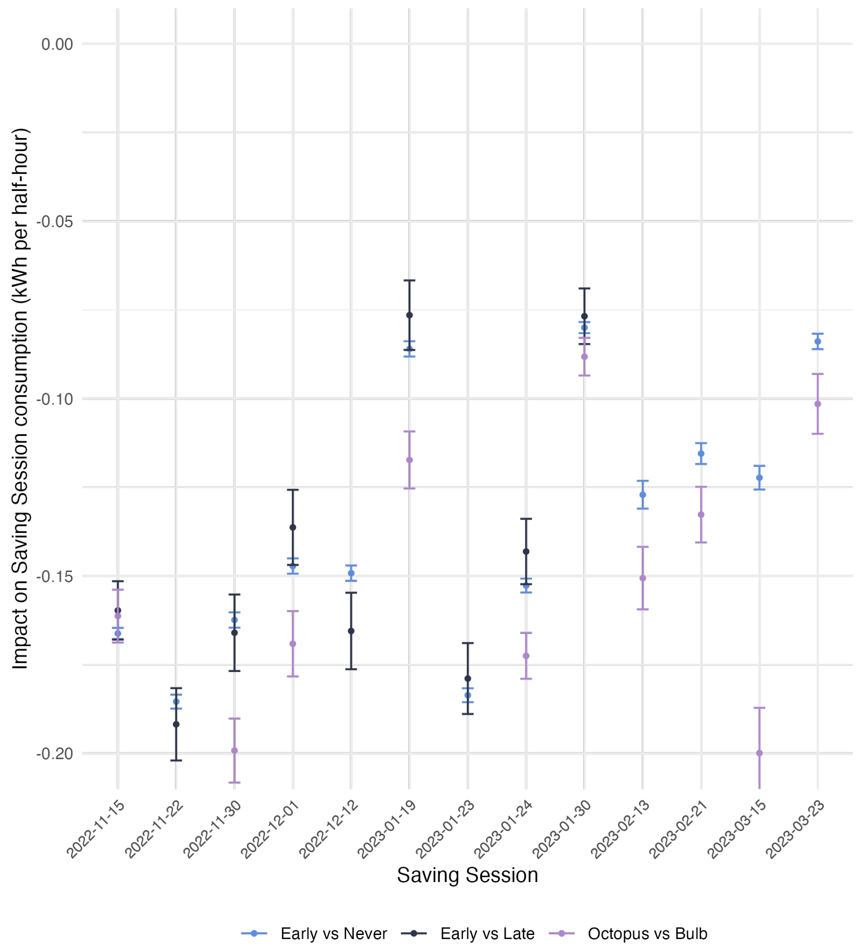
AF.1 DiD Coefficient Plots in kWh

Figure AF.1: Coefficient plot of the impact of sign-up on consumption during Saving Sessions based on our three DiDs.



Note: Coefficient (and 95% confidence intervals) on the difference-in-differences coefficient in our three DiDs for each of the 13 individual Saving Session regressions, where the post-treatment period in each regression is customers' half-hourly consumption during each of the 13 Saving Sessions. In the Octopus versus Bulb DiD, the coefficient is on the local average treatment effect (LATE) of sign-up, a variable equal to 1 if a customer had signed up to Saving Sessions by that Session, else 0. We interpreted these coefficients as the causal impacts of being **signed up** to Saving Sessions by the date of the Session.

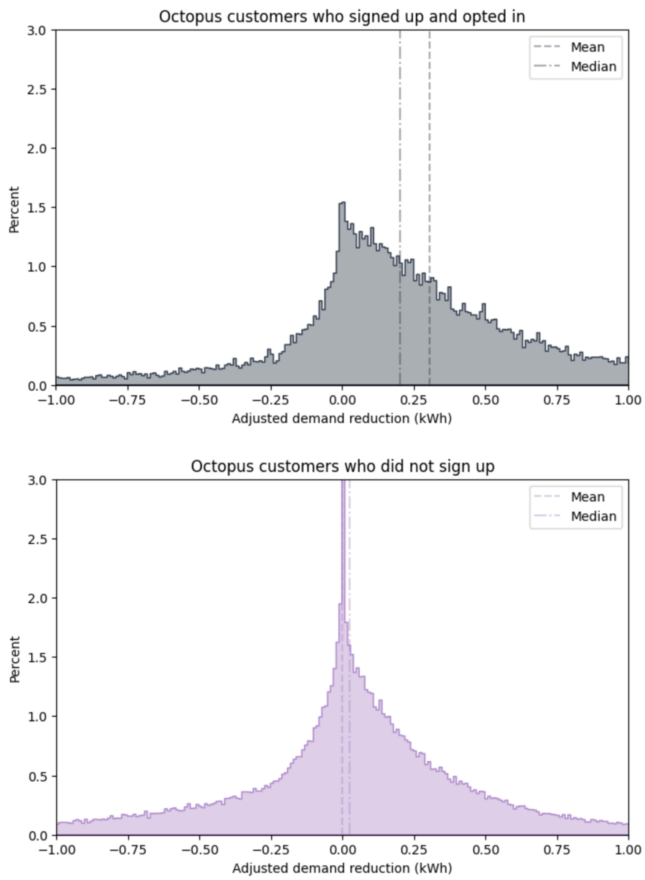
Figure AF.2: Coefficient plot of the impact of opt-in on consumption during Saving Sessions based on our three DiDs.



Note: Coefficient (and 95% confidence intervals) on the difference-in-differences coefficient in our three DiDs for each of the 13 individual Saving Session regressions, where the post-treatment period in each regression is customers' half-hourly consumption during each of the 13 Saving Sessions. In each DiD, the coefficient is on the local average treatment effect (LATE) of opt-in, a variable equal to 1 if a customer opted in to the Session, else 0. We interpreted these coefficients as the causal impacts of *opting in* to Saving Sessions on the date of the Session.

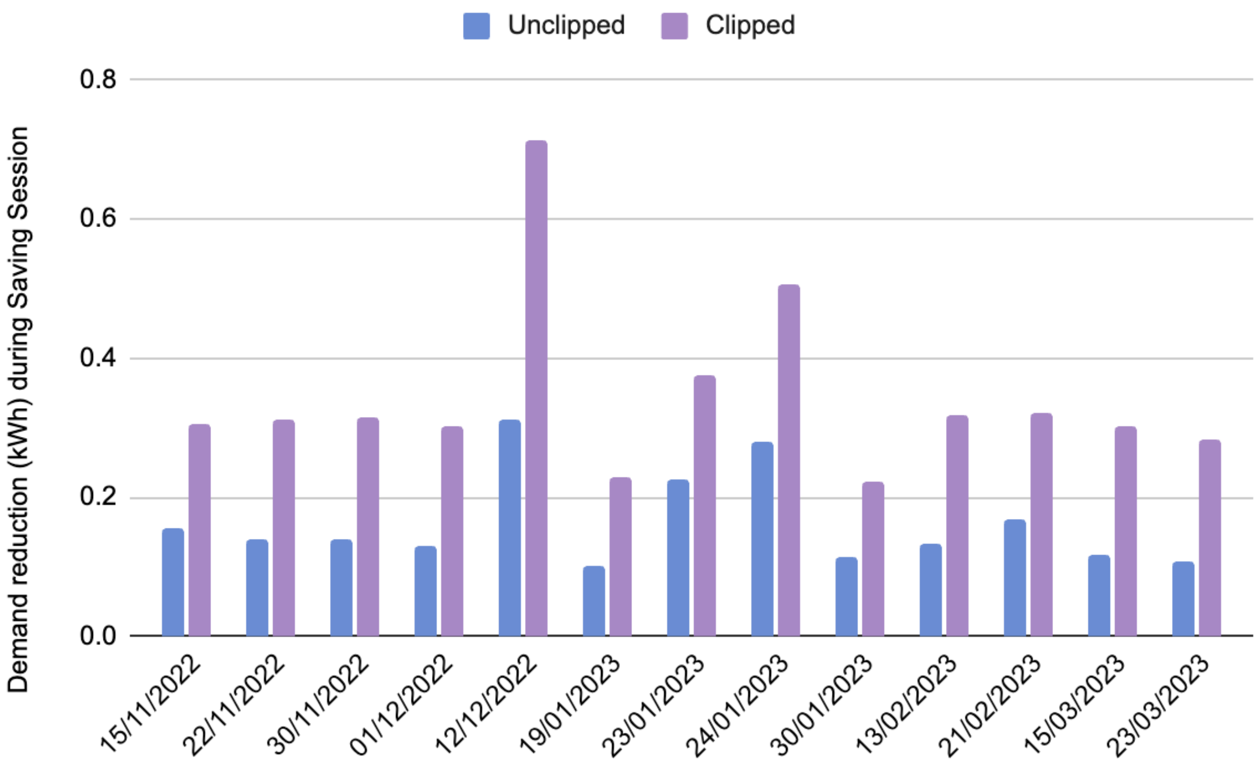
AF.2 Examining Demand Reduction as Measured by NGESO Methodology Depending on Whether Customers Participated

Figure AF.3: Histograms of demand reduction, measured using P376 methodology, by customer participation or non-participation.



Note: Using the "unclipped" P376 methodology prescribed by NGESO results, we see that customers who signed up and opted in for Saving Sessions presented a clear average demand reduction of 0.395 kWh in each half hour of the Saving Sessions. Non-participating customers (customers who signed up but did not opt in, invited Octopus Energy customers who never signed up, and Bulb customers who did not sign up) also show a small demand reduction, according to this methodology, as can be seen by the fact that the histograms for even these customers are not centered around 0. This result suggests that the baselining methodology of choice may overestimate demand reduction.

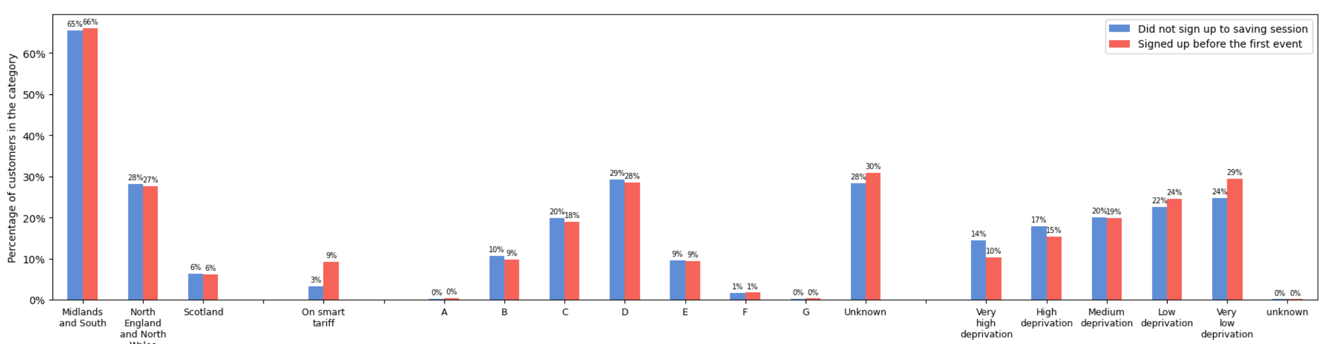
Figure AF.4: Average per-customer "clipped" versus "unclipped" demand reduction, in each Saving Session.



Note: We show clipped versus unclipped demand reduction, across the 13 Saving Sessions Octopus Energy ran between November 2022 and March 2023. The outcome here is demand reduction per customer who had *signed up* to Saving Sessions, regardless of whether they had opted in to a given Session.

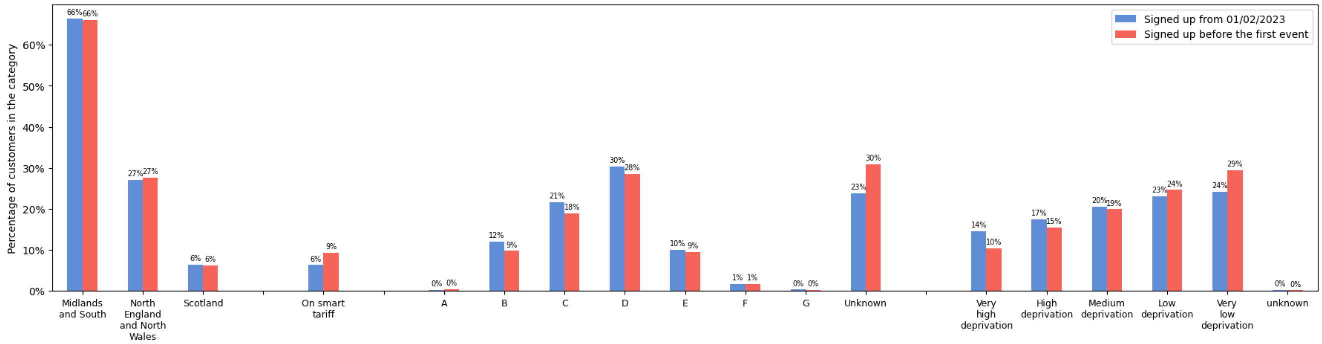
AF.3 Descriptive Statistics of Groups Involved in our Difference-in-differences

Figure AF.5: Region, smart tariff, and EPC grade, by Signed Up Early versus Never customers.



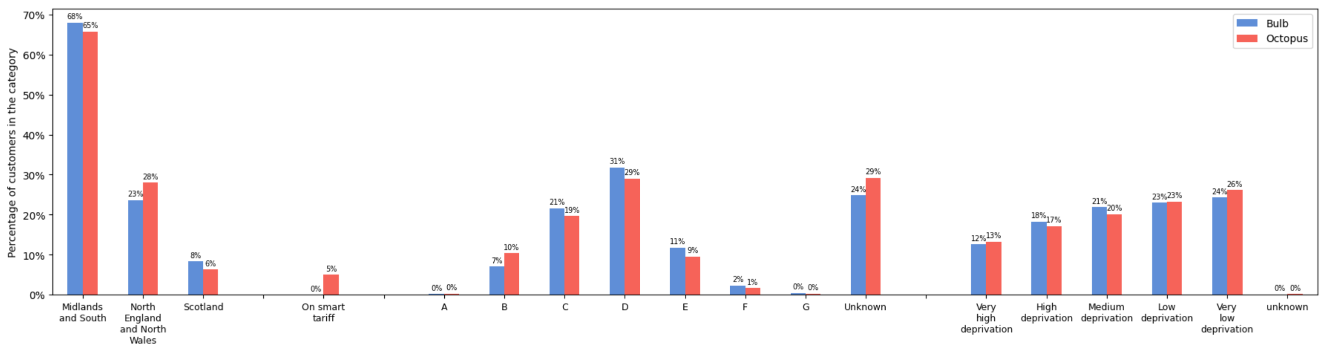
Note: Comparison between Octopus Energy customers that signed up before the first Saving Session (red) and the ones who never signed up (blue) on customers' geographical areas, type of tariff they are on (smart vs not smart), EPC rating, and postcode deprivation level.

Figure AF.6: Region, smart tariff, and EPC grade, by Signed Up Early versus Late.



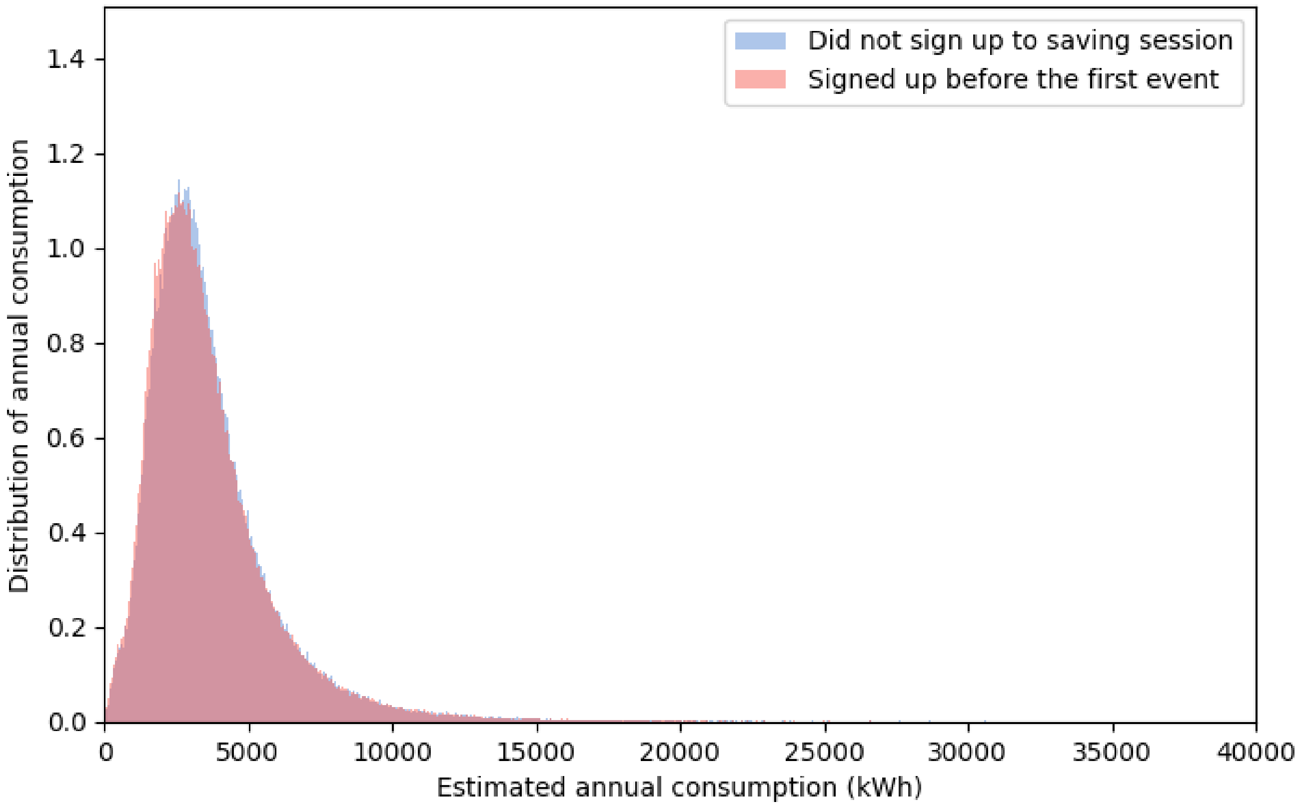
Note: Comparison between Octopus Energy customers that signed up before the first Saving Session (red) and the ones who signed up after 01/02/2023 (blue) on customers' geographical areas, type of tariff they are on (smart vs not smart), EPC rating, and postcode deprivation level.

Figure AF.7: Region, smart tariff, and EPC grade, by Octopus versus Bulb customers.



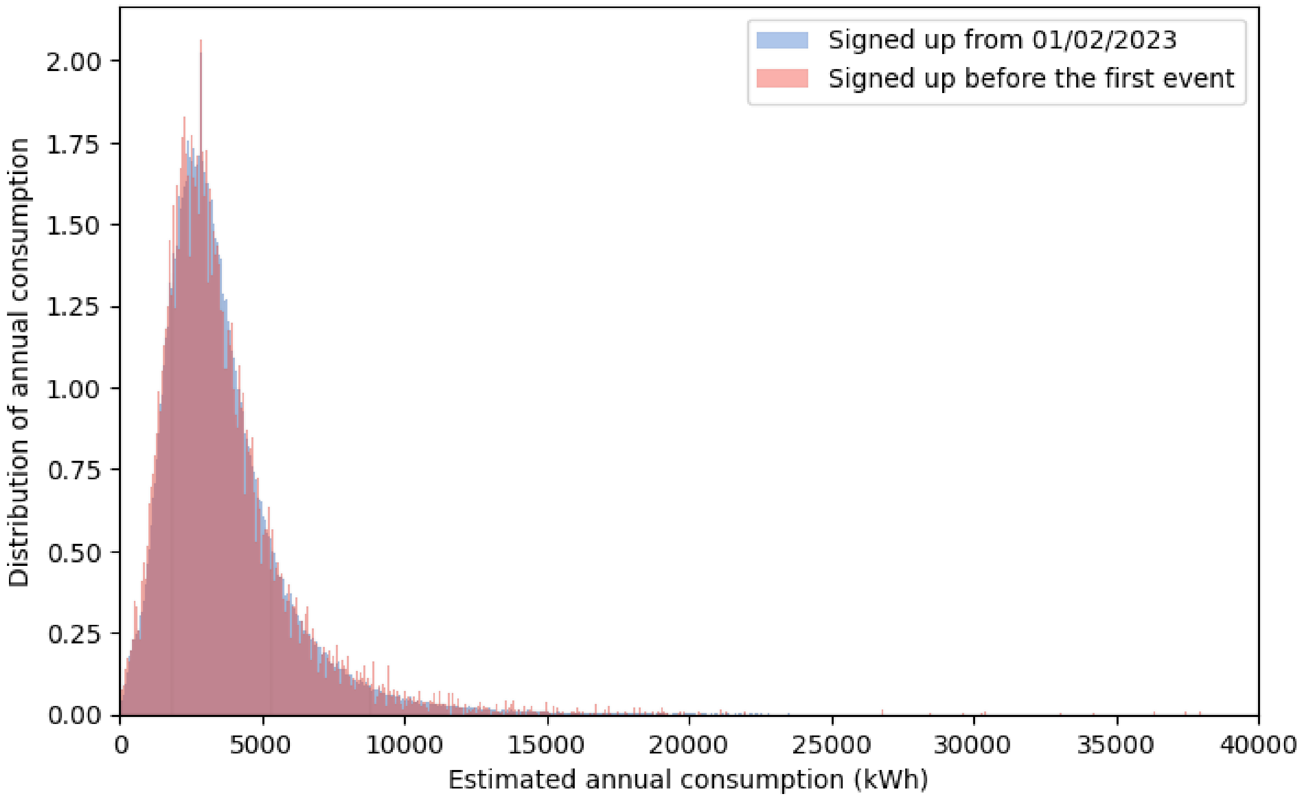
Note: Comparison between Octopus Energy customers (red) and Bulb customers (blue) on customers' geographical areas, type of tariff they are on (smart vs not smart), EPC rating, and postcode deprivation level.

Figure AF.8: Estimated annual electricity consumption distributions: Signed Up Early versus Never.



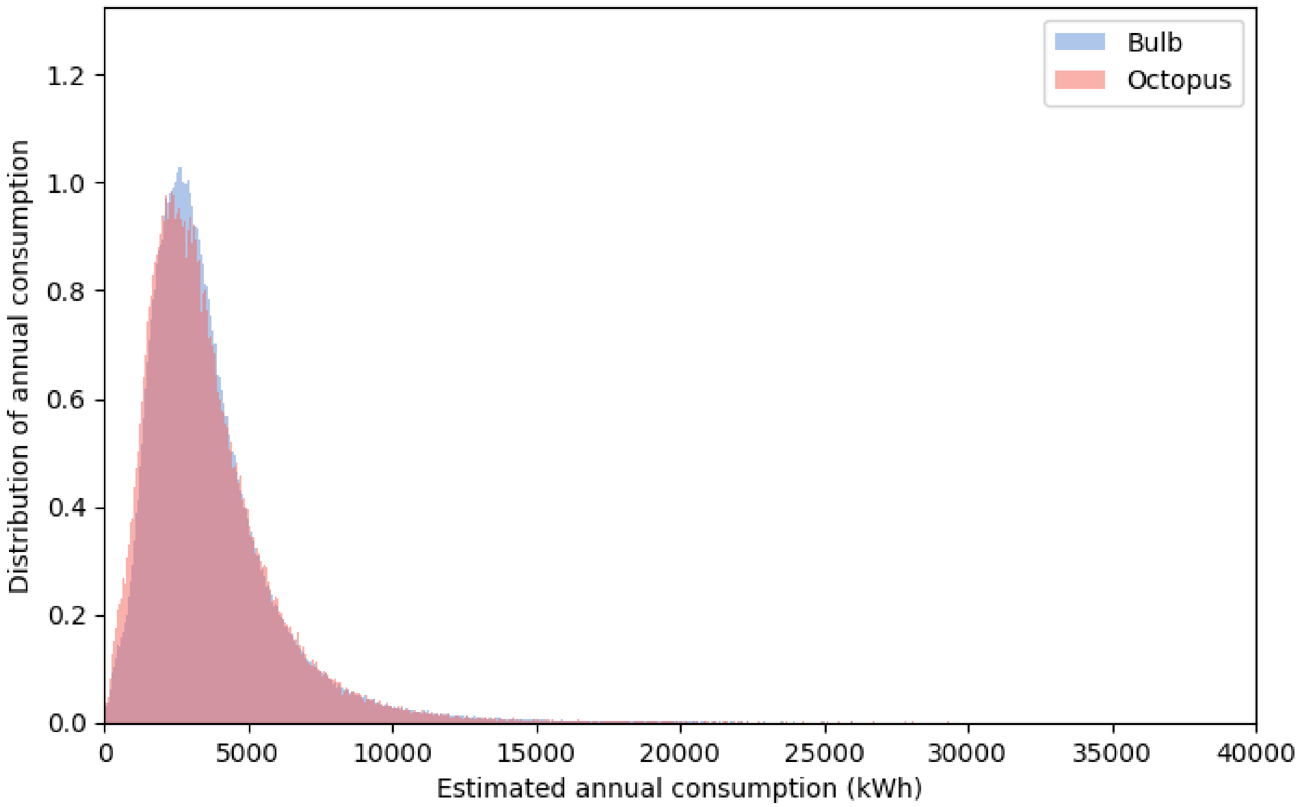
Note: Overlapping histogram of Octopus Energy customers that signed up before the first Saving Session (red; mean 3682.0 kWh; medan 3175.9 kWh) versus those who were invited but never signed up (blue; mean 3643.5 kWh; median 3115.5 kWh) in terms of customers' estimated annual electricity consumption.

Figure AF.9: Estimated annual electricity consumption distributions: Signed Up Early versus Late.



Note: Overlapping histogram of Octopus Energy customers that signed up before the first Saving Session (red; mean 3682.2 kWh; median 3176.0 kWh) versus those who signed up after 01/02/2023 (blue; mean 3720.5 kWh; median 3117.6 kWh) in terms of customers' estimated annual electricity consumption.

Figure AF.10: Estimated annual electricity consumption distributions: Octopus invited versus smart-meter Bulb customers.



Note: Overlapping histogram of Octopus Energy customers invited to Saving Sessions (red; mean 3653.9 kWh; median 3133.2 kWh) versus smart-meter Bulb customers (blue; mean 3568.4 kWh; median 3087.2 kWh) in terms of customers' estimated annual electricity consumption.

AF.4 Flow Diagrams of our DiD Samples

Figure AF.11: Invitation and sign-up flowchart.

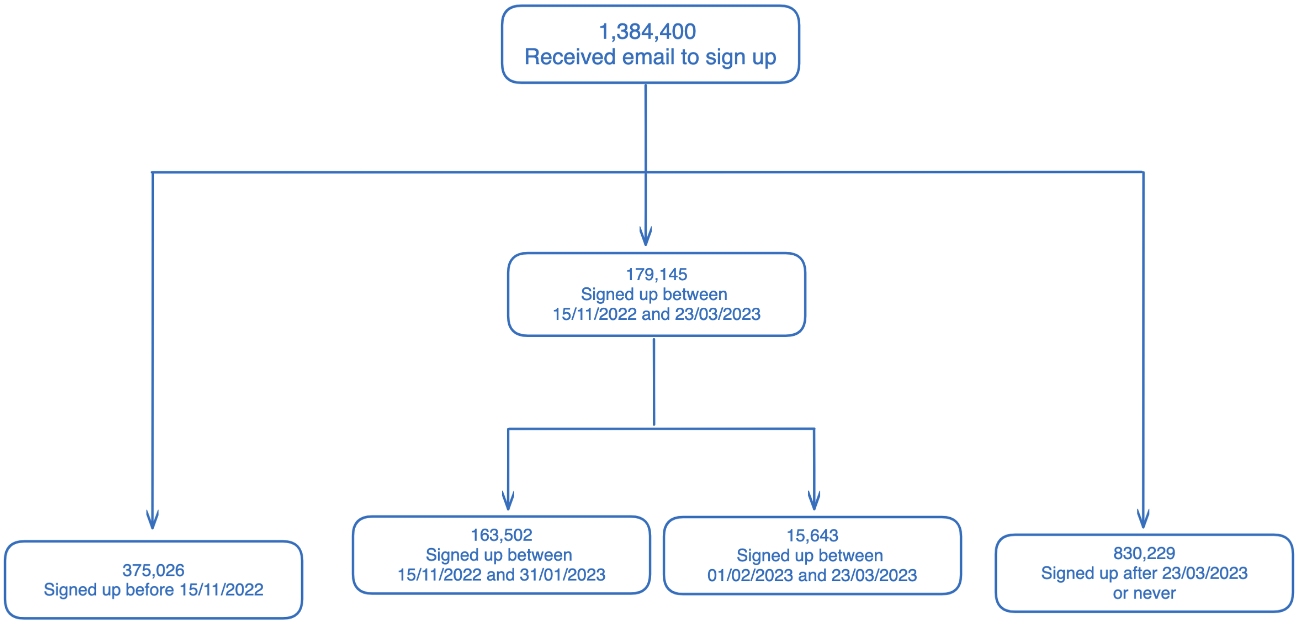


Figure AF.12: Flow diagram showing sample in our Signed Up Early versus Never DiD.

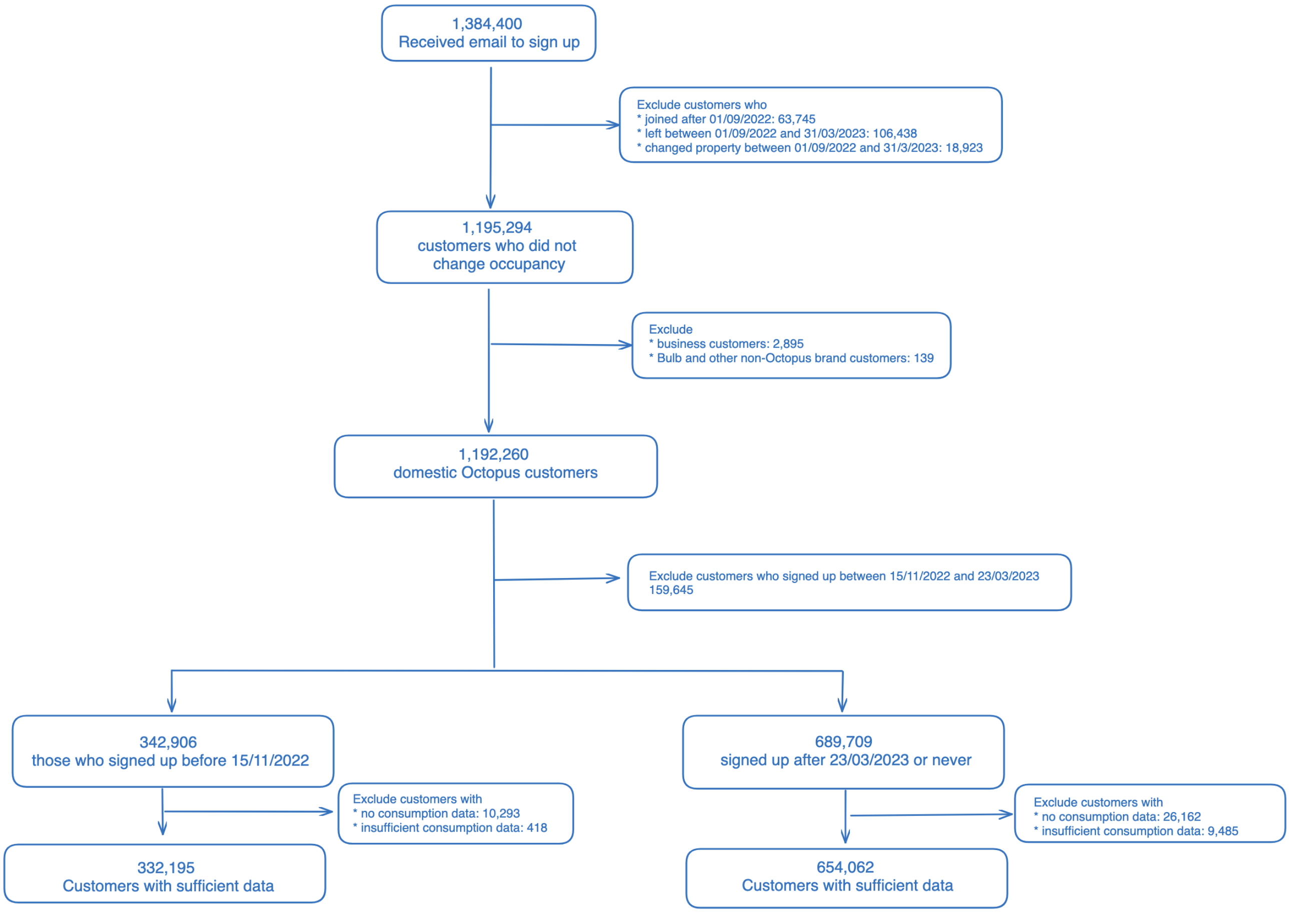


Figure AF.13: Flow diagram showing sample in our Signed Up Early versus Late DiD.

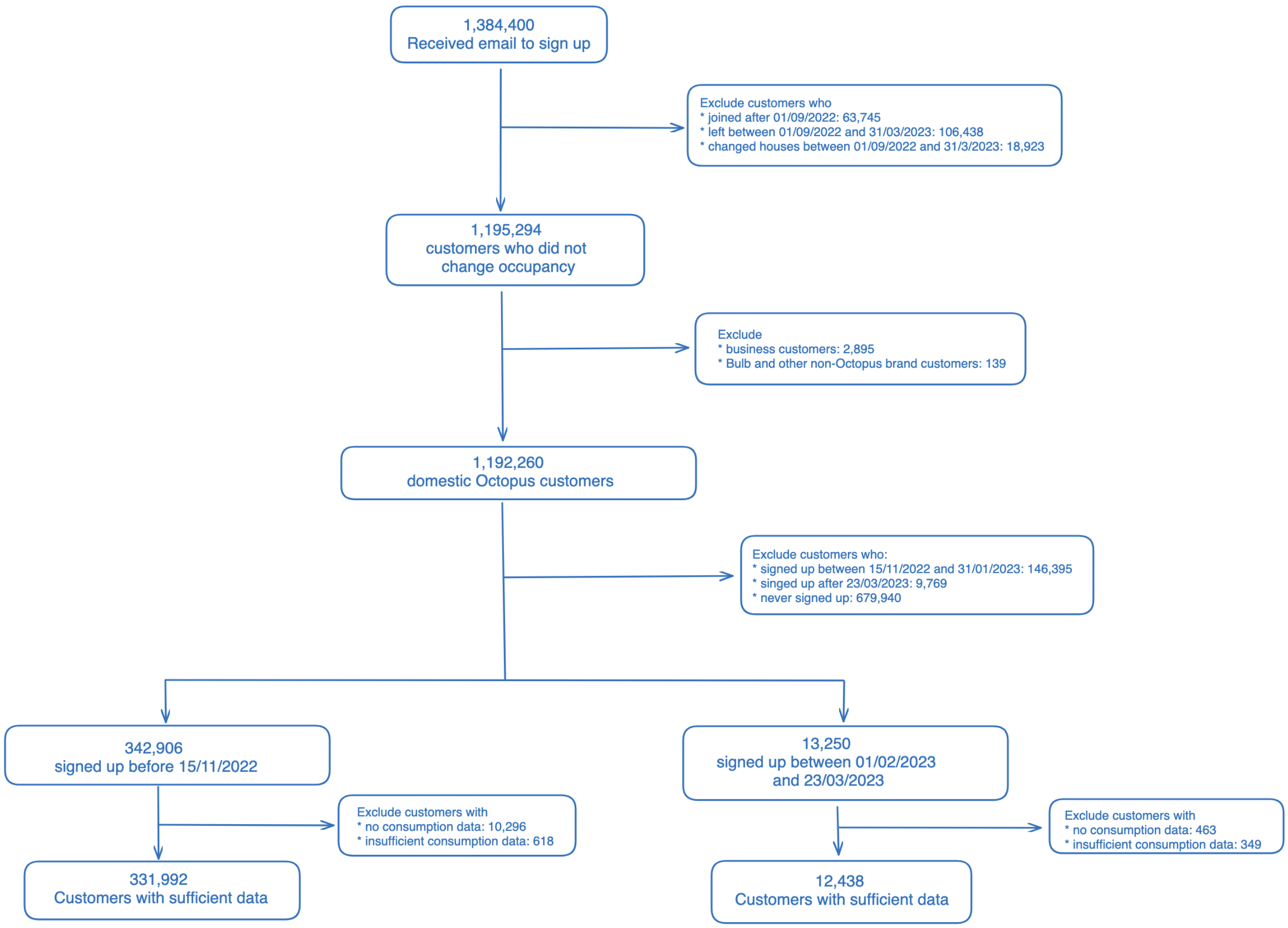
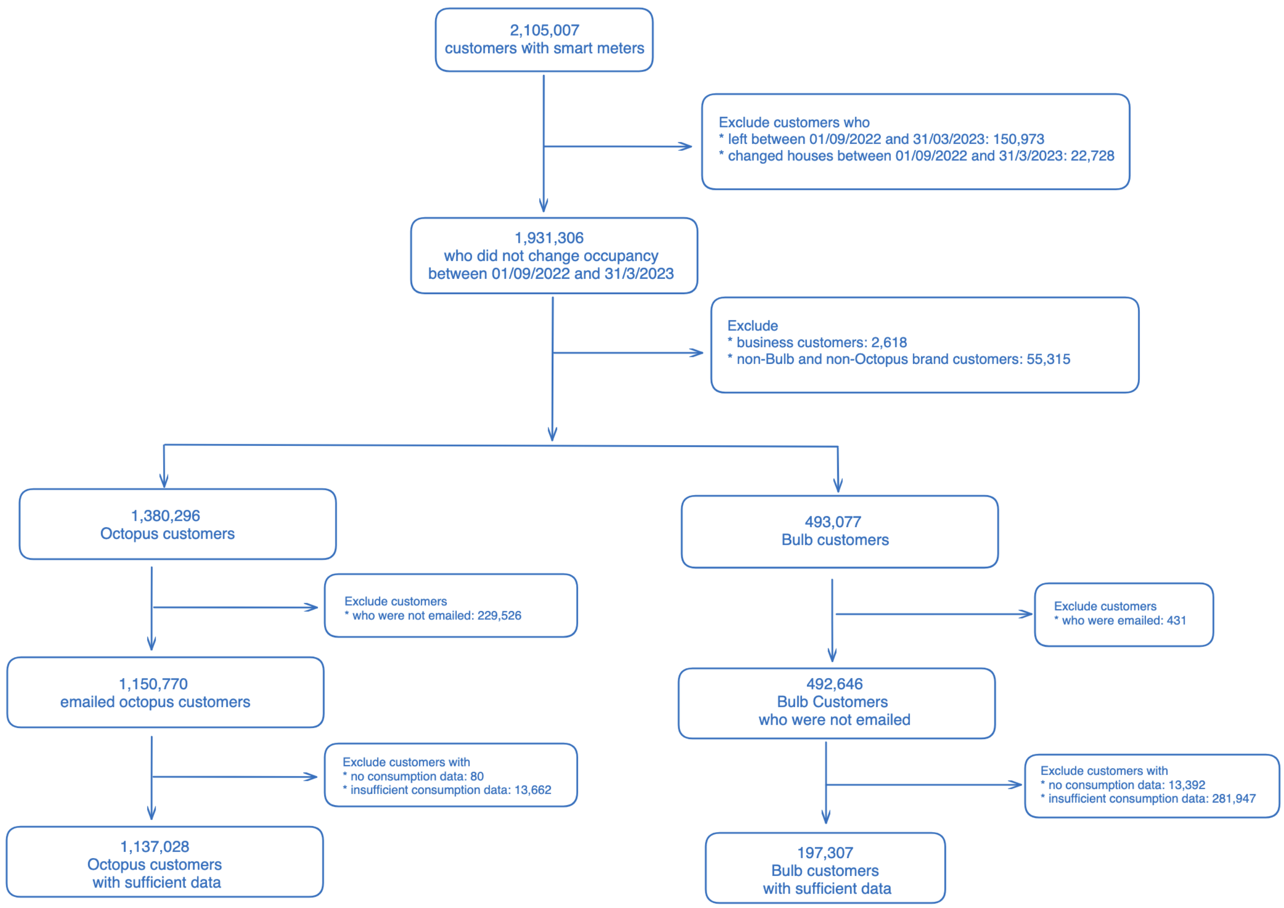
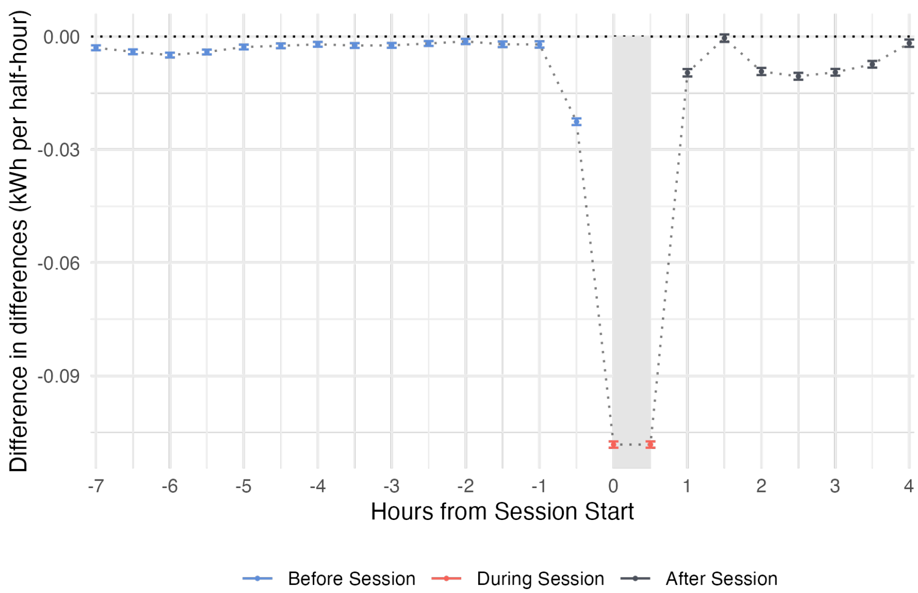


Figure AF.14: Flow diagram showing sample in our Octopus versus Bulb DiD.



AF.5 Event Study of the Six Saving Sessions that Spanned 17:00 to 18:00 or 17:30 to 18:30: Alternative Visualization

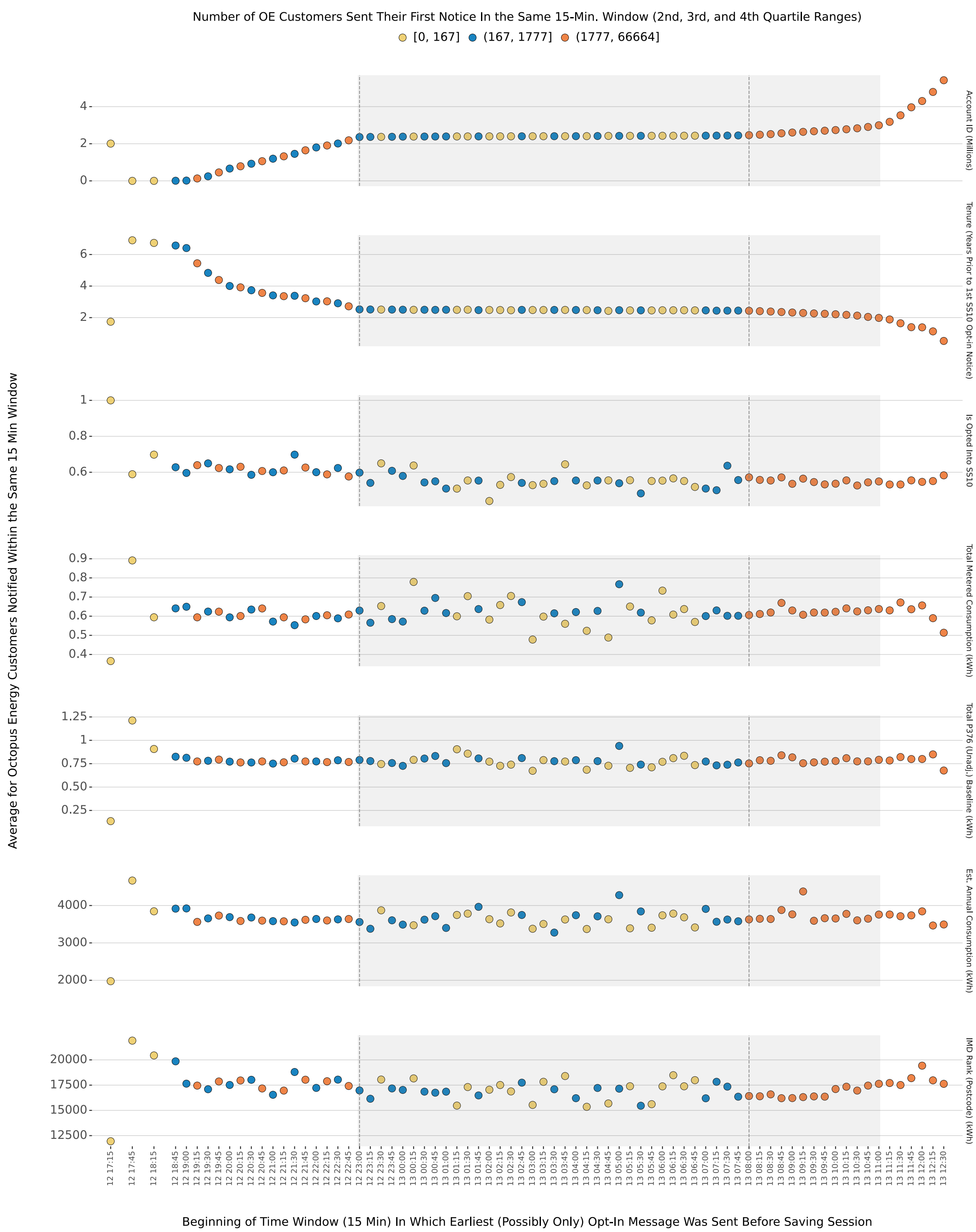
Figure AF.15: Event study of the six Saving Sessions that spanned 17:00 to 18:00 or 17:30 to 18:30.



Note: We show the coefficient on the DiD from a series of regressions using the Signed Up Early versus Never DiD sample. We interpret this coefficient as the demand reduction during the relevant half-hours caused by signing up to Saving Sessions (an ITT effect diluted by incomplete opt-in). The times "0" and "0.5" on the x-axis are from a regression where the post-treatment period is half-hourly consumption during Saving Sessions – two half-hours per Session. All other points on the x-axis are from regressions where the post-treatment period is a single half-hour – X hours before the Session started, if before, or X hours after the Session finished, if after. The whiskers around each point show the 95% confidence interval around the DiD coefficient.

AF.6 RDD: Examination of customer characteristics by time of opt-in notice receipt

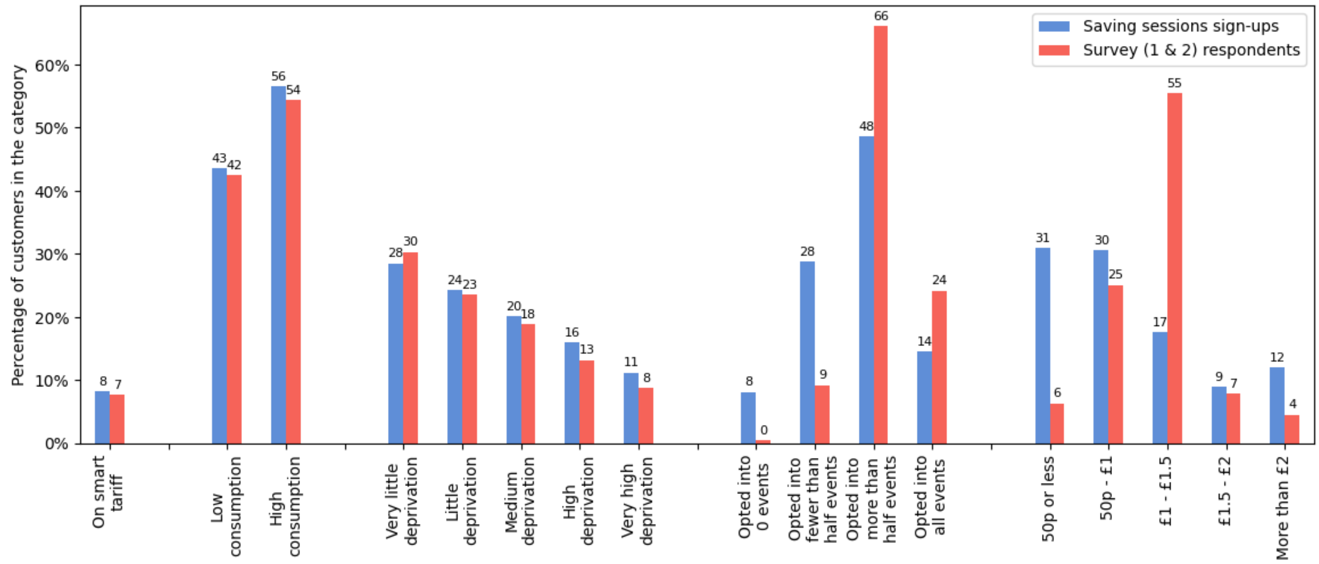
Figure AF.16: "Dot plot" of "binned" averages for response variables and pre-treatment covariates versus temporal running variable.



Note: Averages across Octopus Energy customers who were sent their first (possibly only) opt-in notice for the 10th Saving Session (February 13, 2023) in the same 15-minute window or "bin". Windows wherein zero opt-in notices were sent not shown. The two vertical dashed lines indicate the temporal cutoff for "overnight notices" (left) and treatment (i.e., intraday notice; right; C_{Time}). Shaded region denotes time window corresponding to the range of account IDs used to fit our models of consumption (see Section AI.7).

AF.7 Survey Responses: Analysis and Visualizations

Figure AF.17: Characteristics of Signed-up Saving Sessions Customers and Survey Respondents.



Note: Comparison between Octopus Energy customers that signed up to Saving Sessions (Blue) and Octopus Energy customers who responded to either the survey sent on March 20, 2023 or the survey sent on April 19, 2023 (Red). The two sets of respondents do not overlap. Comparisons are based on the percentage of customers in each of the aforementioned groups that: (a) have a "smart" tariff; (b) have various levels of estimated annual electricity consumption (EAC); (c) live in a postcode with various levels of deprivation; (d) opt into various numbers of Saving Sessions; and (e) that ear various cash values based on their demand reduction during the Sessions they opted into.

Figure AF.18: Responses (April 19, 2023 Survey) to "How interested would you be to keep doing Saving Sessions in the future?".

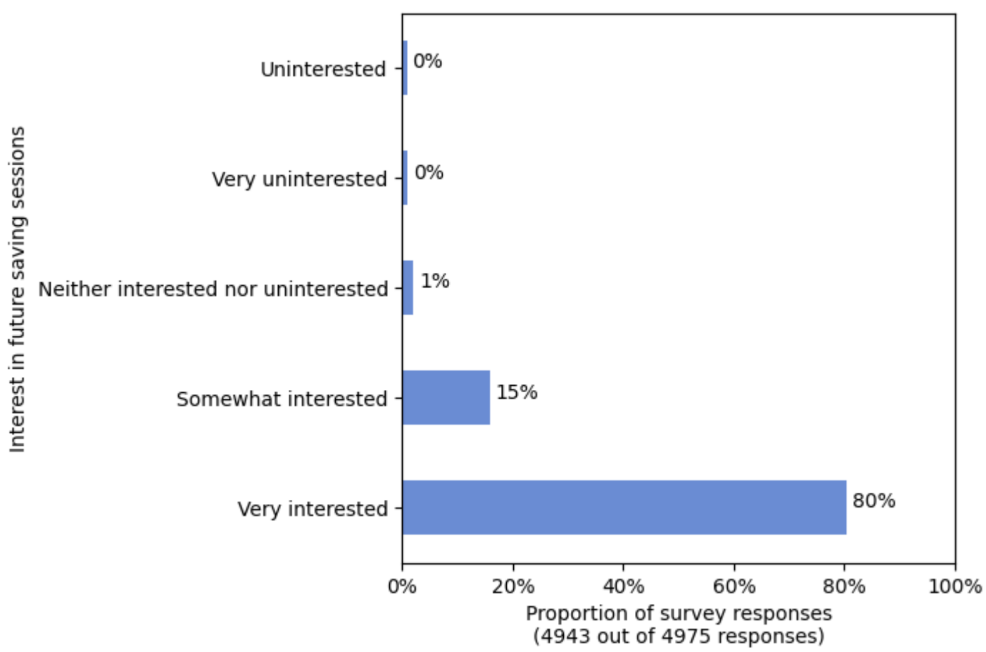


Figure AF.19: Responses (March 20, 2023 Survey) to "If we ran Saving Sessions (or something similar) again next Winter, how many Sessions would you be interested in participating in per week?".

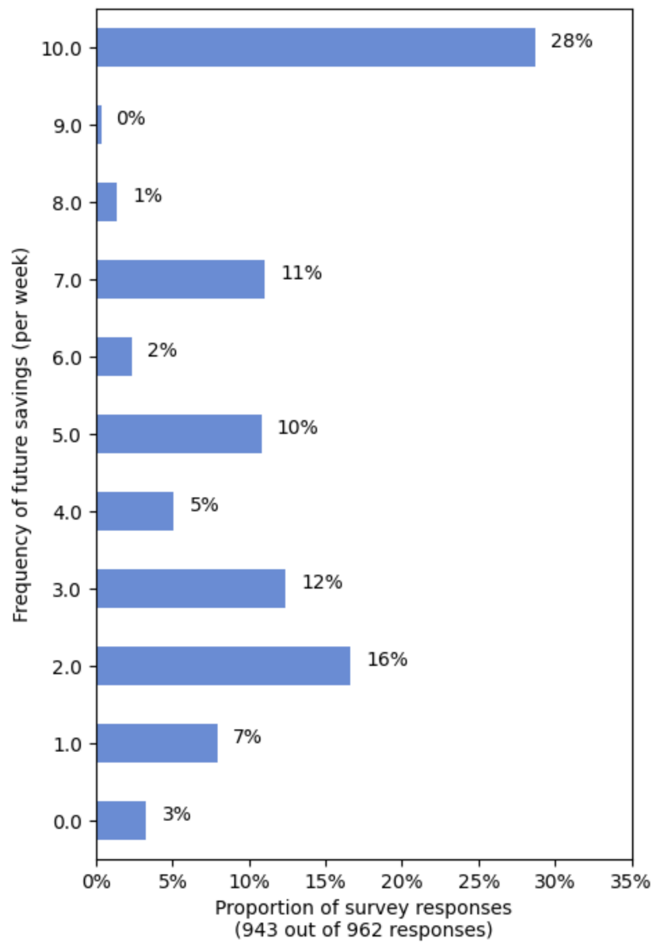
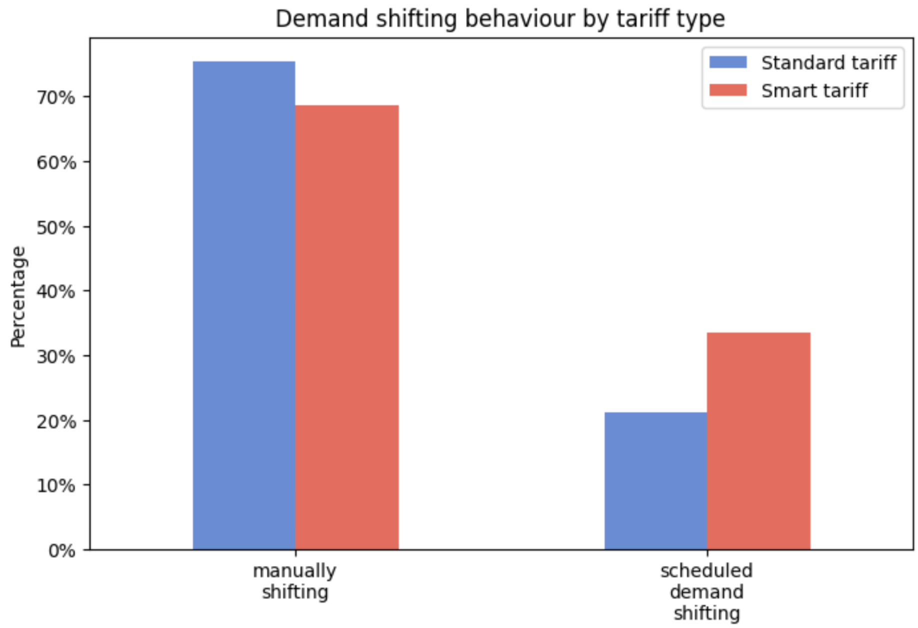
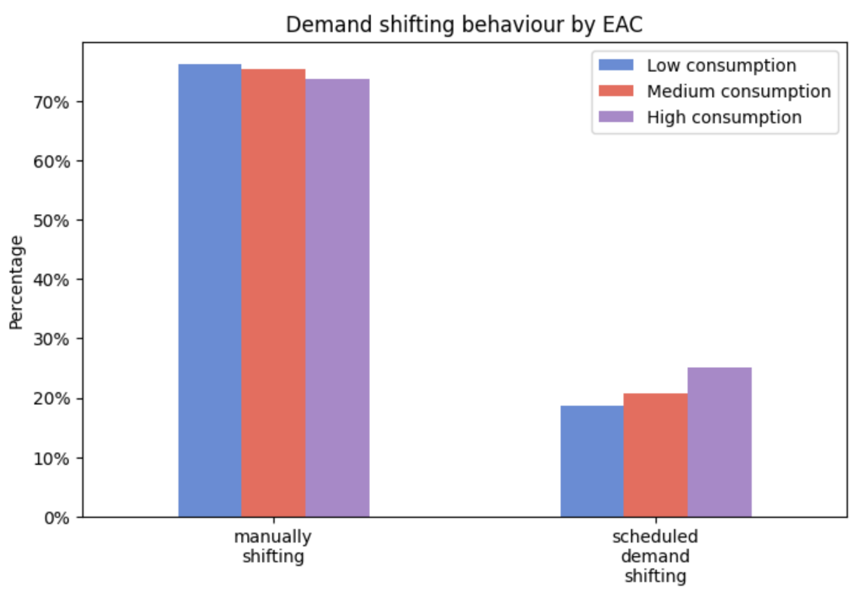


Figure AF.20: Demand shifting behavior by standard vs smart tariff.



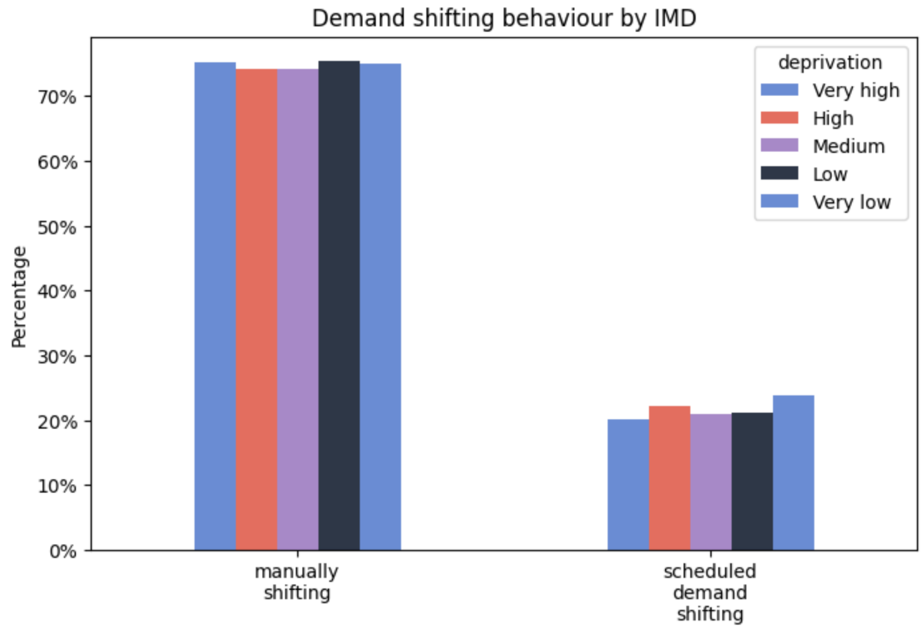
Note: Responses to the question "What best described how you participated" show that most of the consumers for both standard and smart tariffs, 75% and 67% respectively, manually switched off appliances during the session and used them at other times. The remainder for each type of electricity tariff used an automated response, scheduling appliances to come on before or after the session.

Figure AF.21: Demand shifting method by estimated annual electricity consumption level.



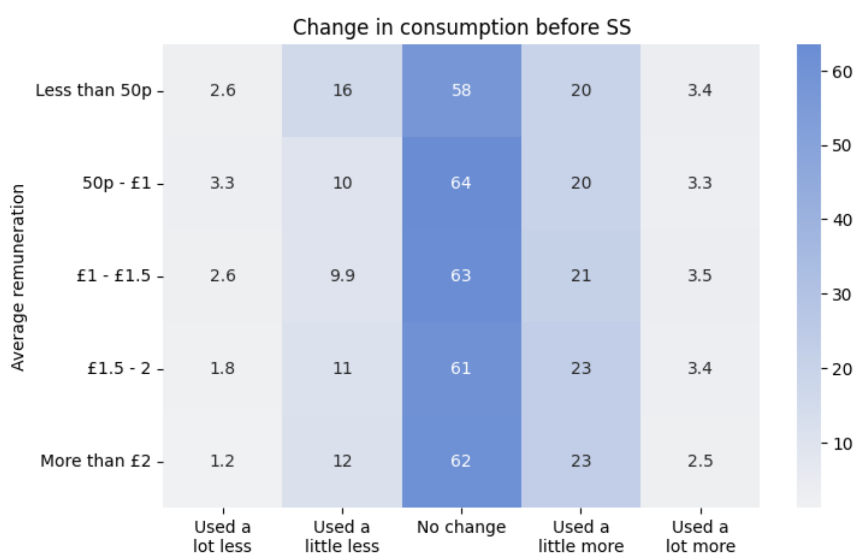
Note: Responses to the question "What best described how you participated" show that most of the consumers, regardless of their level of estimated annual consumption, manually switched off appliances during the session and used them at other times. All three categories are slightly above the 70% level of response. The remainder for each type of estimated annual electricity consumption used an automated response, scheduling appliances to come on before or after the session, being high estimated annual consumption the sub-group of respondents who presented the highest level of automated response, above 25%.

Figure AF.22: Demand shifting behavior by customers' postcode index of multiple deprivation level.



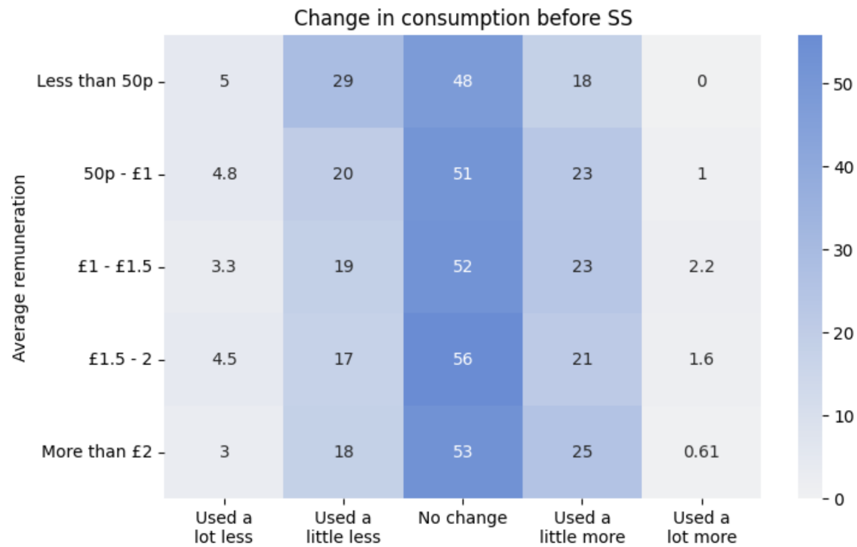
Note: Responses to the question "What best described how you participated" show that most of the consumers for different indices of multiple deprivation, manually switched off appliances during the session and used them at other times. All of these sub-groups presented around 70% of respondents who manually switched off appliances. The remainder for each index of multiple deprivation used an automated response, scheduling appliances to come on before or after the session, being the "Very low" index the one which presents the highest level of automation, slightly above 20%.

Figure AF.23: Change in consumption before Saving Session.



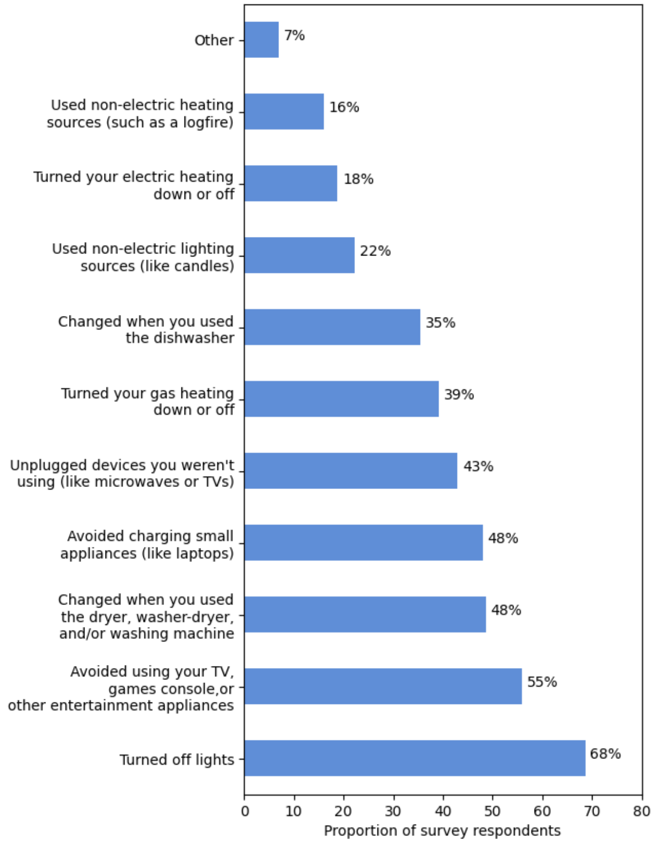
Note: Responses to the question "Did participating in a Saving Session change what you did before the Session?" showed around 60% of the participants did not change their consumption before the session irrespective of the average remuneration they received from the session, i.e. irrespective of the amount of demand reduction they provided. Around 25% of the respondents did use "A little more" or "A lot more" electricity before the session for each of the sub-groups based on the average remuneration. The remainder used "A little less" or "A lot less".

Figure AF.24: Change in consumption after Saving Session.



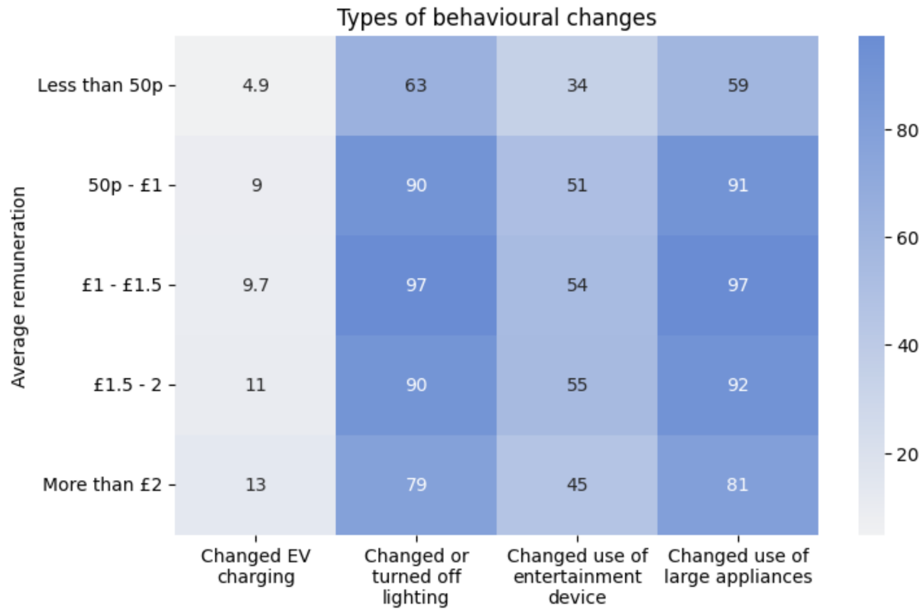
Note: Responses to the question "Did participating in a Saving Session change what you did after the Session?" showed around 50% of the participants did not change their consumption after the session irrespective of the average remuneration they received from the session, i.e. irrespective of the amount of demand reduction they provided. Around 25% of the respondents did use "A little more" or "A lot more" electricity after the session for each of the sub-groups based on the average remuneration except for the lowest level of remuneration (less than 50p) of whom 18% of responses used less energy after the session. The remainder used "A little less" or "A lot less", the lowest average remuneration subgroup (less than 50p) shows again a deviation from the rest of the sub-groups, in this case up to 34% of the respondents used less electricity after the session.

Figure AF.25: Answers to the question "What actions did you do to save energy during Saving Sessions?".



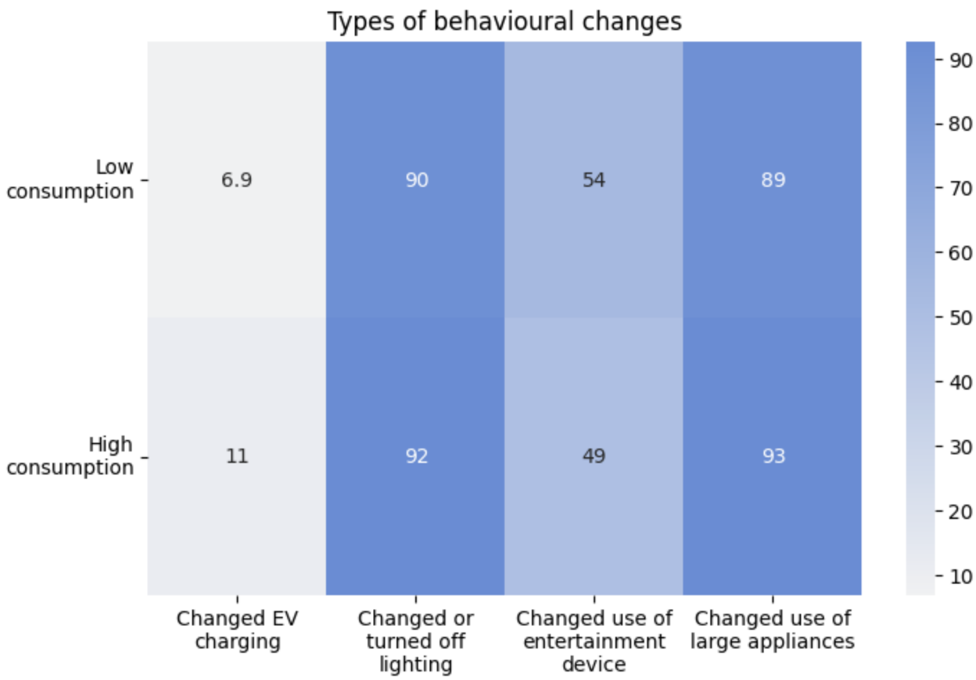
Note: Responses to the question "What actions did you do to save energy during Saving Sessions?" show that participants mostly switched off or plugged off appliances and/or tried to use large appliances at other times of the day to avoid consuming energy during the sessions. The most chosen option was "Turned off lights", chosen by 68% of the respondents. Interestingly, all the answers related to alternative sources of heat or turning off electric heating were less chosen by participants, all of them below 20% of respondents.

Figure AF.26: Behavioral changes by remuneration level.



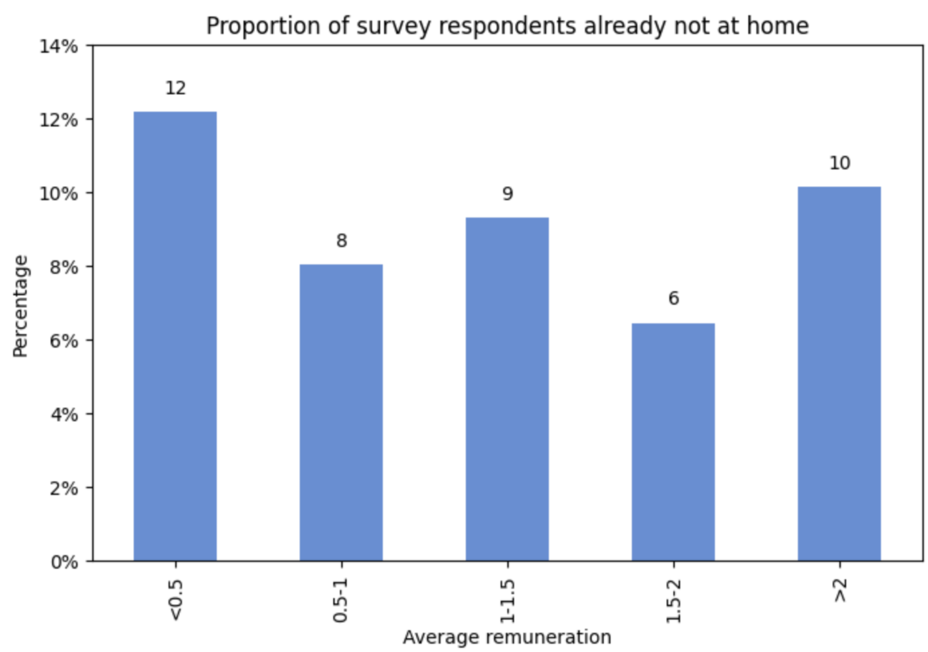
Note: Responses to the question "What actions did you do to save energy during Saving Sessions?" showed interesting results when these are analyzed over different levels of average remuneration. The group of customers that received less than 50p as an average remuneration per session did present lower rates of behavioral changes over all the categories - 1) Changed EV charging, 2) Changed or turned off lights, 3) Changed use of entertainment device, and 4) Change use of large appliances- compared to other customers who got higher incentives. All the different sub-groups based on average remuneration per Saving Session showed a comparable distribution of rates, being option 2 and 4 above the most chosen. There is also an interesting correlation with higher rates of option number one and larger average remunerations per saving session.

Figure AF.27: Behavioral changes by estimated annual electricity consumption level.



Note: Responses to the question "What actions did you do to save energy during Saving Sessions?" showed limited to no variation across different groups of customers based on their average electricity consumption throughout the year. All the categories - 1) Changed EV charging, 2) Changed or turned off lights, 3) Changed use of entertainment device, and 4) Change use of large appliances- show a similar distribution for respondents with low or high consumption. The only noticeable difference is for option 1, preferred for customer with high consumption versus low consumption levels.

Figure AF.28: Proportion of customers who said they tended to be already away from the home during Saving Sessions.



AT Appendix Tables

AT.1 Summary of DiDs' ITT and LATE Results

Table AT.1: Impact (kWh per half-hour) of being invited to sign up.

Saving Session	DiD strategy: Octopus versus Bulb
2022-11-15	-0.0336 (0.0008)
2022-11-22	-0.0437 (0.001)
2022-11-30	-0.0395 (0.0008)
2022-12-01	-0.0328 (0.001)
2022-12-12	-0.0484 (0.0008)
2023-01-19	-0.0231 (0.0008)
2023-01-23	-0.0616 (0.001)
2023-01-24	-0.0477 (0.0008)
2023-01-30	-0.0224 (0.0008)
2023-02-13	-0.0342 (0.001)
2023-02-21	-0.0306 (0.0008)
2023-03-15	-0.0291 (0.001)
2023-03-23	-0.0193 (0.0008)

Note: Coefficient (and standard errors, in parentheses) on the difference-in-differences in our Octopus versus Bulb DiD for each of 13 regressions, where the post-treatment period in each regression is customers' half-hourly consumption during each of the 13 Saving Sessions. We interpreted this coefficient as the causal impact of being *invited to sign up* to Saving Sessions.

Table AT.2: Impact (kWh per half-hour) of signing up.

Saving Session	Signed Up Early versus Never	Signed Up Early versus Late	Octopus versus Bulb
2022-11-15	-0.1172 (0.0005)	-0.1126 (0.0028)	-0.1137 (0.0027)
2022-11-22	-0.1194 (0.0008)	-0.1236 (0.0033)	-0.1424 (0.003)
2022-11-30	-0.1027 (0.0008)	-0.105 (0.0036)	-0.125 (0.0029)
2022-12-01	-0.0906 (0.0008)	-0.0839 (0.0033)	-0.1033 (0.0028)
2022-12-12	-0.1009 (0.0008)	-0.1119 (0.0036)	-0.1481 (0.003)
2023-01-19	-0.0501 (0.0005)	-0.0446 (0.0028)	-0.0664 (0.0023)
2023-01-23	-0.1337 (0.0008)	-0.1302 (0.0038)	-0.1536 (0.0026)
2023-01-24	-0.1063 (0.0008)	-0.0997 (0.0033)	-0.1172 (0.0022)
2023-01-30	-0.0515 (0.0008)	-0.0494 (0.0026)	-0.0541 (0.0017)
2023-02-13	-0.0755 (0.0013)		-0.0838 (0.0025)
2023-02-21	-0.0754 (0.001)		-0.0812 (0.0024)
2023-03-15	-0.0554 (0.0008)		-0.083 (0.0027)
2023-03-23	-0.0503 (0.0008)		-0.0566 (0.0024)

Note: Coefficient (and standard errors, in parentheses) on the difference-in-differences in our three DiDs for each of 13 regressions, where the post-treatment period in each regression is customers' half-hourly consumption during each of the 13 Saving Sessions. In the Octopus versus Bulb DiD, the coefficient is on the local average treatment effect (LATE) of sign-up, a variable equal to 1 if a customer had signed up to Saving Sessions by that Session, else 0. We interpreted these coefficients as the causal impacts of being **signed up** to Saving Sessions by the date of the Session.

Table AT.3: Impact (kWh per half-hour) of opting in.

Saving Session	Signed Up Early versus Never	Signed Up Early versus Late	Octopus versus Bulb
2022-11-15	-0.1662 (0.0008)	-0.1597 (0.0042)	-0.1613 (0.0038)
2022-11-22	-0.1854 (0.001)	-0.1918 (0.0052)	-0.2222 (0.0046)
2022-11-30	-0.1624 (0.0011)	-0.166 (0.0055)	-0.1992 (0.0046)
2022-12-01	-0.1472 (0.0011)	-0.1363 (0.0054)	-0.1691 (0.0047)
2022-12-12	-0.1492 (0.0011)	-0.1655 (0.0055)	-0.2221 (0.0045)
2023-01-19	-0.086 (0.0011)	-0.0765 (0.005)	-0.1173 (0.0041)
2023-01-23	-0.1836 (0.001)	-0.1789 (0.0051)	-0.2174 (0.0036)
2023-01-24	-0.1527 (0.001)	-0.1431 (0.0047)	-0.1725 (0.0033)
2023-01-30	-0.08 (0.0008)	-0.0768 (0.004)	-0.0882 (0.0027)
2023-02-13	-0.1271 (0.002)		-0.1506 (0.0045)
2023-02-21	-0.1155 (0.0015)		-0.1327 (0.004)
2023-03-15	-0.1223 (0.0017)		-0.1999 (0.0065)
2023-03-23	-0.0839 (0.0011)		-0.1015 (0.0043)

Note: Coefficient (and standard errors, in parentheses) on the difference-in-differences in our three DiDs for each of 13 regressions, where the post-treatment period in each regression is customers' half-hourly consumption during each of the 13 Saving Sessions. In each DiD, the coefficient is on the local average treatment effect (LATE) of opt-in, a variable equal to 1 if a customer opted in to the Session, else 0. We interpreted these coefficients as the causal impacts of *opting in* to Saving Sessions on the date of the Session.

Table AT.4: Difference-in-differences results.

	Signed Up Early Never	Signed Up Early Late	Octopus versus Bulb
ITT	-0.0897 (0.0003)	-0.0972 (0.0018)	-0.0361 (0.0005)
LATE on sign-up	–	–	-0.1021 (0.0014)
LATE on opt-in	-0.1425 (0.0006)	-0.1483 (0.0029)	-0.1669 (0.0022)
Mean kwh per half-hour during SS among "control" group	0.376	0.373	0.383
ITT as % of mean	-23.88%	-26.05%	-9.43%
LATE on sign-up as % of mean	–	–	-26.67%
LATE on opt-in as % of mean	-37.94%	-39.74%	-43.59%

Note: The results of the three difference-in-differences (DiDs) with 1) Customers who signed up before the 1st Saving Session versus customers who never signed up, 2) Customers who signed up before the 1st Saving Session versus customers who signed up after the 9th Saving Session and 3) Octopus customers invited to sign up versus smart-meter Bulb customers (who were not invited to participate in the DFS).

AT.2 DiD Regression Outputs: Full Tabular Results for First DiD Strategy

Table AT.5: DiD 1 (Signed Up Early versus Never) results: ITT effect on consumption (kWh per half-hour) during Saving Sessions.

	Coef	Std Err	z-value	$P > \|z\|$	95% CI	
Intercept	0.2509	0.001	455.587	0.000	0.250	0.252
Sign-up	-0.0088	0.000	-29.367	0.000	-0.009	-0.008
Saving Session	0.1989	0.002	119.474	0.000	0.196	0.202
(Saving Session) $*$ Sign-up	-0.0897	0.000	-240.062	0.000	-0.090	-0.089
Average HDDs	-0.1653	0.005	-34.920	0.000	-0.175	-0.156

Note: N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2.

Table AT.6: DiD 1 (Signed Up Early versus Never) results: LATE of opt-in on consumption (kWh per half-hour) during Saving Sessions.

	Coef	Std Err	T-stat	$P > \|T\|$	95% CI	
Intercept	0.2503	0.0005	455.40	0.0000	0.2493	0.2514
Sign-up	-0.0088	0.0003	-29.359	0.0000	-0.0094	-0.0082
Saving Session	0.1972	0.0017	118.70	0.0000	0.1940	0.2005
Average HDDs	-0.1604	0.0047	-33.952	0.0000	-0.1697	-0.1512
Opt in (proportion)	-0.1425	0.0006	-245.66	0.0000	-0.1436	-0.1414

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. The endogenous variable is the proportion of the 13 possible Sessions the customer opted into.

Table AT.7: DiD 1 (Signed Up Early versus Never) results: ITT effect on consumption (kWh per half-hour) in the hour just before Saving Sessions.

	Coef	Std Err	z-value	$P > \|z\|$	95% CI	
Intercept	0.2487	0.001	464.711	0.000	0.248	0.250
Sign-up	-0.0088	0.000	-29.330	0.000	-0.009	-0.008
Just before	0.1350	0.001	94.705	0.000	0.132	0.138
(Just before) * Sign-up	-0.0079	0.000	-22.192	0.000	-0.009	-0.007
Average HDDs	-0.1452	0.005	-31.765	0.000	-0.154	-0.136

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2.

Table AT.8: DiD 1 (Signed Up Early versus Never) results: LATE of opt-in on consumption (kWh per half-hour) in the hour just before Saving Sessions.

	Coef	Std Err	T-stat	P > |T|	95% CI	
Intercept	0.2487	0.0005	464.71	0.0000	0.2476	0.2497
Sign-up	-0.0088	0.0003	-29.329	0.0000	-0.0093	-0.0082
Just before	0.1349	0.0014	94.644	0.0000	0.1321	0.1376
Average HDDs	-0.1448	0.0046	-31.672	0.0000	-0.1538	-0.1359
Opt in (proportion)	-0.0126	0.0006	-22.188	0.0000	-0.0137	-0.0115

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. The endogenous variable is the proportion of the 13 possible Sessions the customer opted into.

Table AT.9: DiD 1 (Signed Up Early versus Never) results: ITT effect on consumption (kWh per half-hour) in the hour just after Saving Sessions.

	Coef	Std Err	z-value	P > |z|	95% CI	
Intercept	0.2515	0.001	459.279	0.000	0.250	0.253
Just after	0.1856	0.002	116.181	0.000	0.182	0.189
Sign-up	-0.0088	0.000	-29.378	0.000	-0.009	-0.008
(Just after) * Sign-up	-0.0089	0.000	-23.447	0.000	-0.010	-0.008
Average HDDs	-0.1712	0.005	-36.403	0.000	-0.180	-0.162

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2.

Table AT.10: DiD 1 (Signed Up Early versus Never) results: LATE of opt-in on consumption (kWh per half-hour) in the hour just after Saving Sessions.

	Coef	Std Err	T-stat	P > |T|	95% CI	
Intercept	0.2514	0.0005	459.33	0.0000	0.2504	0.2525
Sign-up	-0.0088	0.0003	-29.378	0.0000	-0.0094	-0.0082
Just after	0.1854	0.0016	116.13	0.0000	0.1823	0.1885
Average HDDs	-0.1707	0.0047	-36.303	0.0000	-0.1799	-0.1615
Opt in (proportion)	-0.0142	0.0006	-23.450	0.0000	-0.0154	-0.0130

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. The endogenous variable is the proportion of the 13 possible Sessions the customer opted into.

Table AT.11: DiD 1 (Signed Up Early versus Never) results: ITT effect on *daily* consumption (kWh per day) on days of Saving Sessions.

	Coef	Std Err	z-value	P > |z|	[0.025	0.975]
Intercept	12.9356	0.900	14.369	0.000	11.171	14.700
Sign-up	-0.2528	0.439	-0.576	0.565	-1.114	0.608
Saving Session	8.3624	1.674	4.995	0.000	5.081	11.644
(Saving Session) * Sign-up	-0.3769	0.205	-1.835	0.067	-0.780	0.026
Average HDDs	-1.1536	0.334	-3.458	0.001	-1.808	-0.500

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per day during the pre-treatment period (weekdays during October and the first half of November) or post-treatment period (the 13 days with Saving Sessions).

Table AT.12: DiD 1 (Signed Up Early versus Never) results: LATE of opt-in on *daily* consumption (kWh per day) on days of Saving Sessions.

	Coef	Std Err	T-stat	P > |T|	95% CI	
Intercept	12.933	0.9006	14.360	0.0000	11.168	14.698
Sign-up	-0.2528	0.4392	-0.5756	0.5649	-1.1137	0.6081
Saving Session	8.3556	1.6749	4.9887	0.0000	5.0728	11.638
Average HDDs	-1.1523	0.3338	-3.4518	0.0006	-1.8066	-0.4980
Opt in (proportion)	-0.5988	0.3263	-1.8348	0.0665	-1.2384	0.0408

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per day during the pre-treatment period (weekdays during October and the first half of November) or post-treatment period (the 13 days with Saving Sessions). The endogenous variable is the proportion of the 13 possible Sessions the customer opted into.

AT.3 DiD Regression Output: Conditional Average Treatment Effects (CATEs for First DiD Strategy

Table AT.13: CATE for estimated annual electricity consumption (EAC) for DiD 1 (Signed Up Early versus Never).

	Coef	Std Err	z-value	P > |z|	95% CI	
Intercept	0.1562	0.000	324.803	0.000	0.155	0.157
Sign-up	-0.0010	0.000	-4.010	0.000	-0.001	-0.001
High EAC	0.1582	0.000	527.004	0.000	0.158	0.159
Unknown EAC	0.1389	0.008	17.510	0.000	0.123	0.154
Sign-up * (High EAC)	-0.0185	0.000	-38.763	0.000	-0.019	-0.018
Sign-up:Unknown EAC	-0.0369	0.013	-2.933	0.003	-0.062	-0.012
Saving Session	0.1260	0.001	86.183	0.000	0.123	0.129
(Saving Session) * Sign-up	-0.0609	0.000	-159.876	0.000	-0.062	-0.060
(Saving Session) * (High EAC)	0.0908	0.000	208.133	0.000	0.090	0.092
Saving Session:Unknown EAC	0.0863	0.008	11.152	0.000	0.071	0.101
(Saving Session) * Sign-up * (High EAC)	-0.0533	0.001	-76.715	0.000	-0.055	-0.052
(Saving Session) * Sign-up * (Unknown EAC)	-0.0512	0.012	-4.366	0.000	-0.074	-0.028
Average HDDs	-0.1001	0.004	-24.035	0.000	-0.108	-0.092

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. EAC has three levels: high if greater than or equal to 2,900 kWh per year, low if less than 2,900 kWh per year, and unknown if unknown. In this regression, low EAC is the reference category.

Table AT.14: CATE for smart tariff for DiD 1 (Signed Up Early versus Never).

	Coef	Std Err	z-value	P > |z|	95% CI	
Intercept	0.2484	0.001	451.863	0.000	0.247	0.249
Sign-up	-0.0091	0.000	-30.160	0.000	-0.010	-0.008
Smart tariff	0.0606	0.001	42.481	0.000	0.058	0.063
Sign-up * (Smart tariff)	-0.0356	0.002	-20.474	0.000	-0.039	-0.032
Saving Session	0.1957	0.002	117.813	0.000	0.192	0.199
(Saving Session) * Sign-up	-0.0881	0.000	-232.557	0.000	-0.089	-0.087
Saving Session:Smart tariff	0.0522	0.002	28.406	0.000	0.049	0.056
(Saving Session) * Sign-up * (Smart tariff)	-0.0514	0.002	-23.422	0.000	-0.056	-0.047
Average HDDs	-0.1608	0.005	-34.041	0.000	-0.170	-0.152

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. Smart tariff has two levels: True, if the customer's tariff on Nov. 15, 2022 was 'smart', False if any other tariff. (Smart tariffs are special tariffs marketed to customers with low-carbon technologies like electric vehicles, batteries, and heat pumps. Whether a customer is on a smart tariff is a proxy for their engagement with their home's energy consumption and likelihood of having low-carbon technologies like an EV, solar panels, and/or heat pump. In this regression, the reference category is *not* having a smart tariff.

Table AT.15: CATE for index of multiple deprivation quintile for DiD 1 (Signed Up Early versus Never).

	Coef	Std Err	z-value	P > |z|	[0.025	0.975]
Intercept	0.2439	0.001	374.365	0.000	0.243	0.245
Sign-up	-0.0094	0.001	-14.009	0.000	-0.011	-0.008
Very low deprivation	0.0226	0.001	38.462	0.000	0.021	0.024
Low deprivation	0.0092	0.001	15.552	0.000	0.008	0.010
High deprivation	-0.0140	0.001	-24.209	0.000	-0.015	-0.013
Very high deprivation	-0.0203	0.001	-34.179	0.000	-0.021	-0.019
Unknown deprivation level	-0.0236	0.004	-6.603	0.000	-0.031	-0.017
Sign-up * (Very low deprivation)	-0.0067	0.001	-7.398	0.000	-0.008	-0.005
Sign-up * (Low deprivation)	-0.0017	0.001	-1.794	0.073	-0.003	0.000
Sign-up * (High deprivation)	0.0025	0.001	2.644	0.008	0.001	0.004
Sign-up * (Very high deprivation)	0.0024	0.001	2.365	0.018	0.000	0.004
Sign-up * (Unknown deprivation level)	0.0096	0.006	1.594	0.111	-0.002	0.021
Saving Session	0.1887	0.002	108.031	0.000	0.185	0.192
(Saving Session) * Sign-up	-0.0910	0.001	-103.649	0.000	-0.093	-0.089
Saving Session * (Very low deprivation)	0.0111	0.001	15.074	0.000	0.010	0.013
Saving Session * (Low deprivation)	0.0049	0.001	6.506	0.000	0.003	0.006
Saving Session * (High deprivation)	-0.0192	0.001	-25.433	0.000	-0.021	-0.018
Saving Session * (Very high deprivation)	-0.0376	0.001	-50.086	0.000	-0.039	-0.036
Saving Session * (Unknown deprivation level)	-0.0493	0.004	-13.065	0.000	-0.057	-0.042
(Saving Session) * Sign-up * (Very low deprivation)	-0.0149	0.001	-13.029	0.000	-0.017	-0.013
(Saving Session) * Sign-up * (Low deprivation)	-0.0070	0.001	-5.856	0.000	-0.009	-0.005
(Saving Session) * Sign-up * (High deprivation)	0.0127	0.001	10.375	0.000	0.010	0.015
(Saving Session) * Sign-up * (Very high deprivation)	0.0266	0.001	21.038	0.000	0.024	0.029
(Saving Session) * Sign-up * (Unknown deprivation level)	0.0383	0.007	5.337	0.000	0.024	0.052
Average HDDs	-0.1209	0.005	-25.488	0.000	-0.130	-0.112

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. Index of multiple deprivation (IMD) is a measure of relative deprivation for each postcode of the UK. We show IMD *quintiles*: a postcode can be classified as having "very low" deprivation, "low" deprivation, "medium" deprivation, "high" deprivation, or "very high" deprivation. Customer postcode IMD is "unknown" for new postcodes that do not yet have an IMD – these are fewer than 1% of Octopus Energy customers. Thus, IMD has six levels: very low, low, medium, high, very high, and unknown. In this regression, medium (40-60 percentile) IMD is the reference category.

Table AT.16: CATE for EPC letter grade for DiD 1 (Signed Up Early versus Never).

	Coef	Std Err	z-value	P > |z|	[0.025	0.975]
Intercept	0.2478	0.001	408.030	0.000	0.247	0.249
Sign-up	-0.0099	0.001	-18.240	0.000	-0.011	-0.009
EPC: A	-0.0072	0.005	-1.585	0.113	-0.016	0.002
EPC: B	-0.0183	0.001	-29.222	0.000	-0.020	-0.017
EPC: C	-0.0118	0.001	-22.565	0.000	-0.013	-0.011
EPC: E	0.0226	0.001	29.843	0.000	0.021	0.024
EPC: F	0.0623	0.002	28.639	0.000	0.058	0.067
EPC: G	0.0585	0.005	11.866	0.000	0.049	0.068
EPC: Unknown	0.0135	0.000	27.830	0.000	0.013	0.014
Sign-up * (EPC: A)	0.0046	0.007	0.630	0.529	-0.010	0.019
Sign-up * (EPC: B)	0.0049	0.001	4.613	0.000	0.003	0.007
Sign-up * (EPC: C)	0.0036	0.001	4.180	0.000	0.002	0.005
Sign-up * (EPC: E)	-0.0040	0.001	-3.390	0.001	-0.006	-0.002
Sign-up * (EPC: F)	-0.0170	0.003	-5.280	0.000	-0.023	-0.011
Sign-up * (EPC: G)	-0.0079	0.008	-0.980	0.327	-0.024	0.008
Sign-up * (EPC: Unknown)	7.108e-05	0.001	0.093	0.926	-0.001	0.002
Saving Session	0.1995	0.002	116.836	0.000	0.196	0.203
(Saving Session) * Sign-up	-0.0857	0.001	-124.158	0.000	-0.087	-0.084
(Saving Session) * (EPC: A)	0.0983	0.008	12.673	0.000	0.083	0.114
(Saving Session) * (EPC: B)	-0.0238	0.001	-29.995	0.000	-0.025	-0.022
(Saving Session) * (EPC: C)	-0.0126	0.001	-19.118	0.000	-0.014	-0.011
(Saving Session) * (EPC: E)	0.0265	0.001	26.524	0.000	0.025	0.028
(Saving Session) * (EPC: F)	0.0731	0.003	24.891	0.000	0.067	0.079
(Saving Session) * (EPC: G)	0.0951	0.007	13.192	0.000	0.081	0.109
(Saving Session) * (EPC: Unknown)	-0.0016	0.001	-2.666	0.008	-0.003	-0.000
(Saving Session) * Sign-up * (EPC: A)	-0.0395	0.011	-3.516	0.000	-0.061	-0.017
(Saving Session) * Sign-up * (EPC: B)	0.0195	0.001	14.979	0.000	0.017	0.022
(Saving Session) * Sign-up * (EPC: C)	0.0109	0.001	10.226	0.000	0.009	0.013
(Saving Session) * Sign-up * (EPC: E)	-0.0133	0.002	-8.559	0.000	-0.016	-0.010
(Saving Session) * Sign-up * (EPC: F)	-0.0425	0.004	-9.731	0.000	-0.051	-0.034
(Saving Session) * Sign-up * (EPC: G)	-0.0352	0.011	-3.105	0.002	-0.057	-0.013
(Saving Session) * Sign-up * (EPC: Unknown)	-0.0200	0.001	-21.246	0.000	-0.022	-0.018
Average HDDs	-0.1633	0.005	-34.550	0.000	-0.173	-0.154

N=1,972,514 unique observations (986,257 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. In the UK, properties receive Energy Performance Certificate (EPC) letter ratings from A (most efficient) to G (least efficient). They are valid for 10 years. EPCs are needed whenever a property is built, sold, or rented. This means that properties without EPCs are more likely to be owner-occupied (rather than rented) properties that have not been sold in the previous 10 years. EPC thus has seven levels: A through G, and unknown. In this regression, EPC rating "D" is the reference category.

AT.4 DiD Regression Outputs: ITT and LATE Results for Second DiD Strategy

Table AT.17: DiD 2 (Signed Up Early versus Late) results: ITT effect on consumption (kWh per half-hour) during Saving Sessions.

	Coef	Std Err	z-value	P > |z|	95% CI	
Intercept	0.2383	0.002	152.681	0.000	0.235	0.241
Signed up early	-0.0058	0.001	-4.198	0.000	-0.008	-0.003
Saving Session	0.1712	0.003	52.876	0.000	0.165	0.178
Saving Session) * (Signed up early)	-0.0972	0.002	-51.821	0.000	-0.101	-0.094
Average HDDs	-0.0759	0.007	-10.950	0.000	-0.089	-0.062

N=688,860 unique observations (344,430 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. As we show in Figure AF.13, we compare 331,992 Octopus Energy customers who signed up before the first Saving Session to 12,438 Octopus customers who signed up on or after February 1, 2023. We use this specification as our main Signed Up Early versus Late DiD.

Table AT.18: DiD 2 (Signed Up Early versus Late) robustness check: where "late" joiners are defined by joining after the 10th Session.

	Coef	Std Err	z-value	P > |z|	95% CI	
Intercept	0.2394	0.002	121.839	0.000	0.236	0.243
Signed up early	-0.0069	0.002	-3.805	0.000	-0.010	-0.003
Saving Session	0.1724	0.004	47.590	0.000	0.165	0.179
Saving Session) * (Signed up early)	-0.0988	0.002	-40.233	0.000	-0.104	-0.094
Average HDDs	-0.0749	0.007	-10.746	0.000	-0.089	-0.061

N=678,784 unique observations (339,392 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. In this robustness check, the control group of "late" sign-ups is defined by joining on or after February 14, 2023. This date is after the 10th Saving Session on February 13, 2023. This is in contrast to the main Signed Up Early versus Late DiD specification, which defines "late" sign-ups as joining on or after February 1, 2023, after the ninth Session on January 30, 2023. This decreases the size of the "late" group from 12,438 to 7,400, as 5,038 customers joined on or between February 1 and 13, 2023.

Table AT.19: DiD 2 (Signed Up Early versus Late) results: LATE of opt-in on consumption (kWh per half-hour) during Saving Sessions.

	Coef	Std Err	T-stat	P > |T|	95% CI	
Intercept	0.2367	0.0015	152.80	0.0000	0.2336	0.2397
Signed up early	-0.0058	0.0014	-4.1859	0.0000	-0.0085	-0.0031
Saving Session	0.1652	0.0032	52.408	0.0000	0.1590	0.1714
Average HDDs	-0.0601	0.0069	-8.7543	0.0000	-0.0736	-0.0467
Opt in (proportion)	-0.1483	0.0029	-51.887	0.0000	-0.1539	-0.1427

N=688,860 unique observations (344,430 customers, each with one pre and post observation). The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. The endogenous variable is the proportion of the 13 possible Sessions the customer opted into.

Table AT.20: DiD 2 (Signed Up Early versus Late) results: LATE of opt-in on consumption during Saving Sessions where "late" joiners are defined by joining after the 10th Session.

	Coef	Std Err	T-stat	$P > \|T\|$	95% CI	
Intercept	0.2376	0.0020	121.86	0.0000	0.2338	0.2414
Signed up early	-0.0069	0.0018	-3.7964	0.0001	-0.0105	-0.0033
Saving Session	0.1662	0.0035	47.473	0.0000	0.1594	0.1731
Average HDDs	-0.0587	0.0069	-8.4891	0.0000	-0.0722	-0.0451
Opt in (proportion)	-0.1507	0.0037	-40.266	0.0000	-0.1580	-0.1433

N=678,784 unique observations (339,392 customers, each with one pre and post observation). In this robustness check, as in Table AT.18, the control group of "late" sign-ups is defined by joining on or after February 14, 2023. This date is after the 10th Saving Session on February 13, 2023. This is in contrast to the main Signed Up Early versus Late DiD specification, which defines "late" sign-ups as joining on or after February 1, 2023, after the ninth Session on January 30, 2023. This decreases the size of the "late" group from 12,438 to 7,400, as 5,038 customers joined on or between February 1 and 13, 2023. The outcome variable is kWh consumption per half-hour during the pre- or post-treatment period, as described in 3.2. The endogenous variable is the proportion of the 13 possible Sessions the customer opted into.

AT.5 Regression Outputs: Models of Session Participation

Table AT.21: Logistic regression predicting sign-up.

	Coef	Std Err	z-value	P > |z|	[0.025	0.975]
Constant	-0.4729	0.009	-54.370	0.000	-0.490	-0.456
Age band: 18-24	-0.0682	0.008	-9.078	0.000	-0.083	-0.053
Age band: 25-34	-0.0658	0.007	-9.885	0.000	-0.079	-0.053
Age band: 45-54	0.0836	0.005	16.834	0.000	0.074	0.093
Age band: 55-64	0.1045	0.006	16.334	0.000	0.092	0.117
Age band: 65-74	0.0802	0.008	10.068	0.000	0.065	0.096
Age band: 75+	-0.0552	0.009	-5.957	0.000	-0.073	-0.037
Age band: Unknown	-0.1369	nan	nan	nan	nan	nan
Floor area: low	-0.0781	0.006	-12.441	0.000	-0.090	-0.066
Floor area: high	0.1071	0.006	19.455	0.000	0.096	0.118
Floor area: Unknown	0.2425	0.030	8.102	0.000	0.184	0.301
IMD quintile: Very_low	0.1005	0.006	18.159	0.000	0.090	0.111
IMD quintile: Low	0.0483	0.006	8.599	0.000	0.037	0.059
IMD quintile: High	-0.0598	0.007	-8.895	0.000	-0.073	-0.047
IMD quintile: Very_high	-0.1917	0.006	-33.298	0.000	-0.203	-0.180
IMD quintile: Unknown	-0.0087	0.045	-0.192	0.848	-0.097	0.080
EAC: high	-0.0601	0.004	-15.828	0.000	-0.068	-0.053
EAC: Unknown	-0.2727	0.096	-2.831	0.005	-0.461	-0.084
Occupancy type: Multi-occupied_household	-0.2434	0.013	-18.056	0.000	-0.270	-0.217
Occupancy type: Single_adult_household	-0.1425	nan	nan	nan	nan	nan
Occupancy type: Unknown	-0.1369	nan	nan	nan	nan	nan
Rural Urban Classication: rural	0.0527	0.005	11.460	0.000	0.044	0.062
Rural Urban Classication: Unknown	0.0592	0.050	1.189	0.235	-0.038	0.157
On smart tariff	0.9976	0.008	128.186	0.000	0.982	1.013
EPC: A	0.2126	0.033	6.497	0.000	0.148	0.277
EPC: B	-0.0553	0.007	-7.769	0.000	-0.069	-0.041
EPC: C	-0.0016	0.006	-0.281	0.779	-0.013	0.010
EPC: E	-0.0118	0.004	-3.163	0.002	-0.019	-0.004
EPC: F	0.0259	0.015	1.765	0.078	-0.003	0.055
EPC: G	-0.0244	0.032	-0.752	0.452	-0.088	0.039
EPC: Unknown	0.1427	0.009	15.643	0.000	0.125	0.161
DNO: B	0.0682	0.006	11.095	0.000	0.056	0.080
DNO: C	-0.1683	0.009	-18.558	0.000	-0.186	-0.151
DNO: D	0.0501	0.009	5.397	0.000	0.032	0.068
DNO: E	0.0330	0.007	4.446	0.000	0.018	0.048
DNO: F	0.0588	0.008	6.996	0.000	0.042	0.075
DNO: G	0.0159	0.008	1.990	0.047	0.000	0.031
DNO: H	0.1086	0.007	16.147	0.000	0.095	0.122
DNO: J	0.0195	0.008	2.418	0.016	0.004	0.035
DNO: K	0.0728	0.011	6.565	0.000	0.051	0.095
DNO: L	0.1395	0.009	15.522	0.000	0.122	0.157
DNO: M	0.1118	0.007	15.071	0.000	0.097	0.126
DNO: N	0.0222	0.009	2.410	0.016	0.004	0.040
DNO: P	0.0243	0.017	1.391	0.164	-0.010	0.059
DNO: Unknown	-1.5111	0.017	-89.565	0.000	-1.544	-1.478

Note: N=1,384,400 customers invited by Octopus Energy to sign up for Saving Sessions. The outcome variable is whether the customer signed up at any point before or during the Saving Sessions season, which ended on the last Saving Session Mar. 23, 2023. Reference categories are: Age band: 35-44, IMD quintile: Medium, Floor area: medium, EAC: low, Occupancy type: Couple, Rural Urban Classification: urban, *Not* on a smart tariff; EPC: D, and DNO region: A.

Table AT.22: Poisson regression predicting number of opt-ins per customer.

	Coef	Std Err	z-value	$\mathbf{P > \|z\|}$	[0.025	0.975]
Constant	0.9236	0.003	301.463	0.000	0.918	0.930
Age band: 18-24	-0.0640	0.003	-24.088	0.000	-0.069	-0.059
Age band: 25-34	-0.0874	0.002	-35.271	0.000	-0.092	-0.083
Age band: 45-54	0.0612	0.002	29.494	0.000	0.057	0.065
Age band: 55-64	0.0795	0.002	36.539	0.000	0.075	0.084
Age band: 65-74	0.0729	0.002	29.775	0.000	0.068	0.078
Age band: 75+	-0.0328	0.003	-10.836	0.000	-0.039	-0.027
Age band: Unknown	-0.1368	0.003	-45.222	0.000	-0.143	-0.131
Floor area: low	-0.0705	0.002	-28.826	0.000	-0.075	-0.066
Floor area: high	0.0992	0.002	54.474	0.000	0.096	0.103
Floor area: Unknown	0.1962	0.010	20.128	0.000	0.177	0.215
IMD quintile: Very low	0.0759	0.002	40.700	0.000	0.072	0.080
IMD quintile: Low	0.0406	0.002	21.232	0.000	0.037	0.044
IMD quintile: High	-0.0610	0.002	-27.990	0.000	-0.065	-0.057
IMD quintile: Very high	-0.1818	0.003	-71.842	0.000	-0.187	-0.177
IMD quintile: Unknown	0.0002	0.016	0.012	0.991	-0.032	0.032
EAC: high	-0.0766	0.001	-57.530	0.000	-0.079	-0.074
EAC: Unknown	-0.0632	0.032	-1.947	0.051	-0.127	0.000
Occupancy type: Multi-occupied household	-0.2554	0.005	-46.609	0.000	-0.266	-0.245
Occupancy type: Single adult household	-0.1324	0.002	-68.813	0.000	-0.136	-0.129
Occupancy type: Unknown	-0.1368	0.003	-45.222	0.000	-0.143	-0.131
Rural Urban Classication: rural	0.0559	0.002	36.167	0.000	0.053	0.059
Rural Urban Classication: Unknown	0.0958	0.018	5.447	0.000	0.061	0.130
On smart tariff	0.6289	0.002	314.493	0.000	0.625	0.633
EPC: A	0.1957	0.010	20.051	0.000	0.177	0.215
EPC: B	-0.0291	0.002	-12.141	0.000	-0.034	-0.024
EPC: C	-0.0040	0.002	-2.140	0.032	-0.008	-0.000
EPC: E	-0.0146	0.002	-6.106	0.000	-0.019	-0.010
EPC: F	0.0164	0.005	3.314	0.001	0.007	0.026
EPC: G	-0.0551	0.012	-4.768	0.000	-0.078	-0.032
EPC: Unknown	0.1467	0.002	70.684	0.000	0.143	0.151
DNO: B	0.0442	0.003	16.914	0.000	0.039	0.049
DNO: C	-0.1781	0.004	-46.380	0.000	-0.186	-0.171
DNO: D	0.0359	0.003	10.655	0.000	0.029	0.042
DNO: E	0.0279	0.003	10.325	0.000	0.023	0.033
DNO: F	0.0545	0.003	17.914	0.000	0.049	0.060
DNO: G	0.0073	0.003	2.564	0.010	0.002	0.013
DNO: H	0.0722	0.003	28.215	0.000	0.067	0.077
DNO: J	0.0092	0.003	3.263	0.001	0.004	0.015
DNO: K	0.0335	0.004	8.669	0.000	0.026	0.041
DNO: L	0.1179	0.003	38.417	0.000	0.112	0.124
DNO: M	0.0987	0.003	34.713	0.000	0.093	0.104
DNO: N	0.0064	0.003	1.940	0.052	-6.64×10^{-5}	0.013
DNO: P	0.0362	0.006	6.179	0.000	0.025	0.048
DNO: Unknown	-1.4090	0.008	-186.872	0.000	-1.424	-1.394

Note: N=969,080 (this is 70% of the 1,384,400 customers who received an email to sign up to Saving Sessions; 30% were held back to check model accuracy). The outcome variable is how many events the customer opted into (of a possible 13). Reference categories are: Age band: 35-44, IMD quintile: Medium, Floor area: medium, EAC: low, Occupancy type: Couple, Rural Urban Classification: urban, *Not* on a smart tariff; EPC: D, and DNO region: A.

Table AT.23: Logistic regression predicting opt-in.

	Coef	Std Err	z-value	$P > \|z\|$	[0.025	0.975]
Constant	0.4173	0.009	46.564	0.000	0.400	0.435
Age band: 18-24	-0.0541	0.008	-6.993	0.000	-0.069	-0.039
Age band: 25-34	-0.0520	0.007	-7.237	0.000	-0.066	-0.038
Age band: 45-54	0.0231	0.006	3.779	0.000	0.011	0.035
Age band: 55-64	0.0537	0.006	8.417	0.000	0.041	0.066
Age band: 65-74	0.0628	0.007	8.815	0.000	0.049	0.077
Age band: 75+	0.0057	0.009	0.650	0.516	-0.012	0.023
Age band: Unknown	-0.0681	0.034	-1.983	0.047	-0.135	-0.001
Floor area: low	0.0005	0.007	0.065	0.948	-0.013	0.014
Floor area: high	0.0662	0.005	12.367	0.000	0.056	0.077
Floor area: Unknown	0.1422	0.029	4.902	0.000	0.085	0.199
IMD quintile: Very_low	0.0419	0.005	7.676	0.000	0.031	0.053
IMD quintile: Low	0.0243	0.006	4.347	0.000	0.013	0.035
IMD quintile: High	-0.0457	0.006	-7.261	0.000	-0.058	-0.033
IMD quintile: Very_high	-0.1190	0.007	-16.505	0.000	-0.133	-0.105
IMD quintile: Unknown	0.0024	0.049	0.049	0.961	-0.094	0.099
EAC: high	-0.1034	0.004	-26.743	0.000	-0.111	-0.096
EAC: Unknown	0.0851	0.110	0.776	0.438	-0.130	0.300
Occupancy type: Multi-occupied_household	-0.1469	0.015	-9.728	0.000	-0.177	-0.117
Occupancy type: Single_adult_household	-0.0762	0.005	-13.867	0.000	-0.087	-0.065
Occupancy type: Unknown	-0.0681	0.033	-2.077	0.038	-0.132	-0.004
Rural Urban Classication: rural	0.0557	0.005	12.214	0.000	0.047	0.065
Rural Urban Classication: Unknown	0.0725	0.053	1.374	0.170	-0.031	0.176
On smart tariff	0.3800	0.006	59.253	0.000	0.367	0.393
EPC: A	0.2481	0.031	7.912	0.000	0.187	0.310
EPC: B	0.0049	0.007	0.691	0.489	-0.009	0.019
EPC: C	0.0164	0.006	2.971	0.003	0.006	0.027
EPC: E	-0.0281	0.007	-4.068	0.000	-0.042	-0.015
EPC: F	-0.0028	0.014	-0.197	0.844	-0.031	0.025
EPC: G	-0.0568	0.033	-1.716	0.086	-0.122	0.008
EPC: Unknown	0.0472	0.005	8.866	0.000	0.037	0.058
DNO: B	0.0175	0.008	2.284	0.022	0.002	0.033
DNO: C	-0.1266	0.011	-11.744	0.000	-0.148	-0.105
DNO: D	-0.0095	0.010	-0.980	0.327	-0.029	0.010
DNO: E	-0.0019	0.008	-0.245	0.807	-0.017	0.013
DNO: F	0.0257	0.009	2.886	0.004	0.008	0.043
DNO: G	-0.0200	0.008	-2.420	0.016	-0.036	-0.004
DNO: H	0.0190	0.007	2.545	0.011	0.004	0.034
DNO: J	-0.0208	0.008	-2.547	0.011	-0.037	-0.005
DNO: K	-0.0437	0.011	-3.916	0.000	-0.066	-0.022
DNO: L	0.0769	0.009	8.515	0.000	0.059	0.095
DNO: M	0.0489	0.008	5.866	0.000	0.033	0.065
DNO: N	-0.0355	0.010	-3.661	0.000	-0.055	-0.016
DNO: P	0.0681	0.018	3.850	0.000	0.033	0.103
DNO: Unknown	0.0480	0.021	2.309	0.021	0.007	0.089

Note: N=6,057,082 customer * event combinations. We use 543,235 customers who signed up before the final Saving Session, each of whom has between one and 13 observations – potential Sessions they chose whether or not to opt into – depending on when during the Saving Sessions season they signed up. The outcome variable is whether the customer opted into the event. Reference categories are: Age band: 35-44, IMD quintile: Medium, Floor area: medium, EAC: low, Occupancy type: Couple, Rural Urban Classification: urban, *Not* on a smart tariff; EPC: D, and DNO region: A.

Table AT.24: Logistic regression predicting customer opt-in for the 8th Saving Session on Jan. 24, 2023.

	Coef	Std Err	z-value	P > |z|	[0.025	0.975]
Constant	-0.8211	0.018	-46.084	0.000	-0.856	-0.786
Age band: 18-24	0.0033	0.015	0.224	0.823	-0.026	0.032
Age band: 25-34	-0.0436	0.013	-3.301	0.001	-0.070	-0.018
Age band: 45-54	0.0247	0.012	2.128	0.033	0.002	0.047
Age band: 55-64	0.0981	0.013	7.625	0.000	0.073	0.123
Age band: 65-74	0.1598	0.014	11.083	0.000	0.132	0.188
Age band: 75+	0.1172	0.017	6.919	0.000	0.084	0.150
Age band: Unknown	-0.0589	0.109	-0.541	0.589	-0.273	0.155
Floor area: low	0.0136	0.013	1.043	0.297	-0.012	0.039
Floor area: high	0.0987	0.010	9.836	0.000	0.079	0.118
Floor area: Unknown	0.1955	0.056	3.501	0.000	0.086	0.305
IMD quintile: Very low	0.0698	0.011	6.610	0.000	0.049	0.091
IMD quintile: Low	0.0294	0.011	2.730	0.006	0.008	0.051
IMD quintile: High	-0.0485	0.012	-4.076	0.000	-0.072	-0.025
IMD quintile: Very high	-0.1200	0.013	-8.955	0.000	-0.146	-0.094
IMD quintile: Unknown	-0.0668	0.088	-0.755	0.450	-0.240	0.107
EAC: high	-0.0815	0.007	-10.998	0.000	-0.096	-0.067
EAC: Unknown	-0.0059	0.191	-0.031	0.975	-0.381	0.369
Occupancy type: Multi-occupied household	-0.1415	0.028	-5.075	0.000	-0.196	-0.087
Occupancy type: Single adult household	-0.0803	0.010	-7.683	0.000	-0.101	-0.060
Occupancy type: Unknown	-0.0589	0.097	-0.606	0.544	-0.250	0.132
Rural Urban Classication: rural	0.0624	0.009	7.045	0.000	0.045	0.080
Rural Urban Classication: Unknown	0.1713	0.095	1.798	0.072	-0.015	0.358
On smart tariff	0.3336	0.012	26.699	0.000	0.309	0.358
EPC: A	0.1793	0.060	2.973	0.003	0.061	0.297
EPC: B	0.0070	0.013	0.531	0.595	-0.019	0.033
EPC: C	0.0178	0.010	1.721	0.085	-0.002	0.038
EPC: E	-0.0185	0.013	-1.404	0.160	-0.044	0.007
EPC: F	-0.0254	0.028	-0.918	0.358	-0.080	0.029
EPC: G	-0.0115	0.064	-0.180	0.857	-0.137	0.114
EPC: Unknown	0.0771	0.010	7.413	0.000	0.057	0.097
DNO: B	-0.0115	0.015	-0.787	0.431	-0.040	0.017
DNO: C	-0.0427	0.020	-2.090	0.037	-0.083	-0.003
DNO: D	0.0014	0.019	0.073	0.942	-0.036	0.038
DNO: E	-0.0199	0.015	-1.320	0.187	-0.049	0.010
DNO: F	-0.0437	0.017	-2.569	0.010	-0.077	-0.010
DNO: G	-0.0352	0.016	-2.229	0.026	-0.066	-0.004
DNO: H	0.0595	0.014	4.108	0.000	0.031	0.088
DNO: J	0.0086	0.016	0.546	0.585	-0.022	0.039
DNO: K	0.0171	0.020	0.854	0.393	-0.022	0.056
DNO: L	0.0186	0.018	1.025	0.305	-0.017	0.054
DNO: M	0.0376	0.019	2.013	0.044	0.001	0.074
DNO: N	0.0104	0.017	0.605	0.545	-0.024	0.044
DNO: P	-0.0266	0.015	-1.762	0.078	-0.056	0.003
DNO: Unknown	0.0002	0.025	0.009	0.993	-0.049	0.050
is_opted_into_23jan_event	2.2606	0.007	318.654	0.000	2.247	2.274

Note: N=507,493 customers signed up by the 7th Saving Session on Jan. 23, 2023. The outcome variable is whether the customer opted into the Saving Session on Jan. 24, 2023. Reference categories are: Age band: 35-44, IMD quintile: Medium, Floor area: medium, EAC: low, Occupancy type: Couple, Rural Urban Classification: urban, *Not* on a smart tariff; EPC: D, and DNO region: A.

AT.6 RDD and Field Trial: Descriptive Statistics

Table AT.25: Global descriptive statistics for correlates used for the regression discontinuity and field trial.

	N_{Missing}	Binary	N_{True}	Mean	Std. Deviation
Total Metered Consumption (kWh)	3,378	No	—	0.618	2.704
Opted Into Session [Ref Not Opted In]	0	No	—	0.570	0.495
Intraday Opt-in Notice [Ref Day-ahead Notice]	0	Yes	209,312	—	—
Intraday Opt-in Notice (07:59:59, 09:00:00] [Ref Day-ahead Notice]	0	Yes	14,672	—	—
Intraday Opt-in Notice (09:00:00, 10:00:00] [Ref Day-ahead Notice]	0	Yes	15,632	—	—
Intraday Opt-in Notice (10:00:00, 11:00:00] [Ref Day-ahead Notice]	0	Yes	32,189	—	—
Intraday Opt-in Notice (11:00:00, 12:00:00] [Ref Day-ahead Notice]	0	Yes	173,279	—	—
Intraday Opt-in Notice (12:00:00, 13:00:00] [Ref Day-ahead Notice]	0	Yes	164,447	—	—
Account ID	0	No	—	2991160.533	1500449.293
Avg. Historical in-Session Consumption (kWh)	952	No	—	0.700	8.068
Avg. Historical Session Participation	894	No	—	0.527	0.332
Total P376 (Unadjusted) Baseline Consumption (kWh)	0	No	—	0.788	6.904
Estimated Annual Consumption (kWh)	63,097	No	—	3671.738	4798.974
Index of Multiple Deprivation (IMD) Rank (Postcode)	80,001	No	—	17821.793	9893.760
Tenure with Octopus (Years Prior to 1st SS10 Opt-in Notice)	62,084	No	—	2.312	1.189
Business Entity [Ref Non-Business Entity]	0	Yes	561	—	—
Has Non-Octopus Product [Ref Octopus Product]	0	Yes	10,404	—	—
Has Smart Tariff [Ref Non-Smart Tariff]	0	Yes	41,609	—	—
DNO Region: North [Ref Midlands & South]	0	Yes	150,552	—	—
DNO Region: Scotland [Ref Midlands & South]	0	Yes	34,439	—	—

(**a**) Regression Discontinuity Design ($N = 609{,}531$).

	N_{Missing}	Binary	N_{True}	Mean	Std. Deviation
Total Metered Consumption (kWh)	12,567	No	—	0.654	0.784
Opted Into Session [Ref Not Opted In]	0	No	—	0.445	0.497
Intraday Notice + Day-ahead Email [Ref Intraday Only]	0	Yes	19,182	—	—
Intraday Notice + Intraday SMS + £1.25 Incentive [Ref Intraday Only]	0	Yes	4,472	—	—
Intraday Notice + Intraday SMS + £1.25 Incentive Assigned	0	Yes	19,220	—	—
Avg. Historical in-Session Consumption (kWh)	8,678	No	—	0.667	3.825
Avg. Historical Session Participation	8,569	No	—	0.527	0.330
Total P376 (Unadjusted) Baseline Consumption (kWh)	0	No	—	0.763	2.178
Estimated Annual Consumption (kWh)	73,607	No	—	3660.404	4691.735
Index of Multiple Deprivation (IMD) Rank (Postcode)	91,651	No	—	17756.542	9893.062
Business Entity [Ref Non-Business Entity]	0	Yes	592	—	—
Has Non-Octopus Product [Ref Octopus Product]	0	Yes	30,925	—	—
Has Smart Tariff [Ref Non-Smart Tariff]	0	Yes	41,098	—	—
DNO Region: North [Ref Midlands & South]	0	Yes	157,854	—	—
DNO Region: Scotland [Ref Midlands & South]	0	Yes	37,231	—	—

(**b**) Randomized Field Trial ($N = 650{,}809$).

Note: We use the binary indicator *Opted Into Session* as the response variable for our linear probability models and here we report its the standard deviation as $\sigma_p = \sqrt{\mu_p \times (1 - \mu_p)}$. We incorporate continuous variables into our models for our regression discontinuity design and our field trial by first constructing Z-scores using our full dataset and then filtering our data based on missing values and, in the case of our RDD, a bandwidth h. Z-scores were constructed by subtracting the global mean in the sample and dividing by the global standard deviation. For our regression discontinuity, we construct *Account ID* relative to the cutoff for treatment which we then re-scale to millions prior to model fitting. Moreover, we use three measures of historical energy usage — i.e., *Total P376 (Unadjusted) Baseline (kWh)*, *Estimated Annual Consumption (kWh)*, and *Average Historical in-Session Consumption (kWh)*. The first measure is an unweighted average of consumption during the same half-hour of the day during the ten most-recent working days as governed by the the P376 amendment to Great Britain's electricity balancing and settlement code. The second measure is Octopus Energy's predicted customer consumption based on meter readings over the past year. The third measure is a customer's average consumption across all Saving Sessions, regardless of Session opt-in but after DFS sign-up, *prior to* February 13, 2023 (regression discontinuity design) or March 15, 2023 (field trial). We also use a measure of the average (i.e., the proportion) of Saving Sessions that a customer opted into after DFS sign-up *prior to* February 13, 2023 (regression discontinuity design) or March 15, 2023 (field trial). For the deprivation index, more deprived areas have *lower* postcode ranks.

AT.7 RDD Regression Output: Complete Results

Table AT.26: Models of total consumption (kWh) during the 10th Saving Session (regression discontinuity design).

MSE-Optimal Bandwidth ($h_{\text{Left}}, h_{\text{Right}}$)	$(\mathbf{h_L}, \mathbf{h_R})$	$(\mathbf{h_L}, \mathbf{h_R})$	$(\mathbf{h_{L\div 1.5}}, \mathbf{h_{R\times 1.5}})$	$(\mathbf{h_{L\div 2}}, \mathbf{h_{R\times 2}})$	$\mathbf{h_L}$ **Only**
$\hat{\beta}$ **Intercept**	0.566 (0.024)	0.588 (0.020)	0.532 (0.048)	0.653 (0.034)	0.613 (0.006)
$\hat{\beta}$ **Intraday Opt-in Notice**	0.060 (0.025)	0.042 (0.021)	0.109 (0.049)	-0.021 (0.034)	—
$\hat{\beta}$ **Intraday Opt-in Notice (07:59:59, 09:00:00]**	—	—	—	—	0.009 (0.008)
$\hat{\beta}$ **Intraday Opt-in Notice (09:00:00, 10:00:00]**	—	—	—	—	0.012 (0.008)
$\hat{\beta}$ **Intraday Opt-in Notice (10:00:00, 11:00:00]**	—	—	—	—	0.026 (0.007)
$\hat{\beta}$ **Intraday Opt-in Notice (11:00:00, 12:00:00]**	—	—	—	—	0.037 (0.007)
$\hat{\beta}$ **Intraday Opt-in Notice (12:00:00, 13:00:00]**	—	—	—	—	0.056 (0.008)
$\hat{\beta}$ **Account ID (Millions) [Relative to Cutoff]**	-0.201 (0.138)	-0.247 (0.117)	-0.724 (0.371)	0.306 (0.295)	—
$\hat{\beta}$ **Intraday Opt-in Notice** $\times$ **Account ID**	0.210 (0.139)	0.252 (0.114)	0.679 (0.370)	-0.345 (0.294)	—
$\hat{\beta}$ **Avg. Hist. Sess. Consumption (kWh) [Z-Score]**	—	2.804 (0.125)	2.678 (0.101)	3.896 (0.112)	3.988 (0.075)
$\hat{\beta}$ **Avg. Hist. Sess. Participation [Z-Score]**	—	-0.023 (0.003)	-0.024 (0.002)	-0.006 (0.002)	-0.004 (0.002)
$\hat{\beta}$ **Total P376 Baseline (kWh) [Z-Score]**	—	2.305 (0.113)	2.389 (0.092)	4.649 (0.078)	4.541 (0.071)
$\hat{\beta}$ **Est. Annual Consumption (kWh) [Z-Score]**	—	0.059 (0.010)	0.071 (0.012)	-0.243 (0.014)	-0.219 (0.010)
$\hat{\beta}$ **IMD Rank (Postcode) [Z-Score]**	—	-0.007 (0.003)	-0.003 (0.002)	-0.013 (0.002)	-0.015 (0.001)
$\hat{\beta}$ **Tenure with Octopus (Years) [Z-Score]**	—	-0.047 (0.039)	-0.082 (0.037)	-0.087 (0.036)	0.030 (0.005)
$\hat{\beta}$ **Business Entity [Ref Non-Business]**	—	0.455 (0.274)	0.217 (0.250)	-0.097 (0.174)	-0.365 (0.090)
$\hat{\beta}$ **Has Non-Octopus Product [Ref Octo. Prod.]**	—	-0.091 (0.034)	-0.047 (0.038)	-0.027 (0.033)	-0.055 (0.018)
$\hat{\beta}$ **Has Smart Tariff [Ref Non-Smart Tariff]**	—	-0.064 (0.010)	-0.080 (0.010)	0.007 (0.010)	0.010 (0.006)
$\hat{\beta}$ **DNO Region: North [Ref Midlands & South]**	—	0.013 (0.005)	0.016 (0.005)	0.030 (0.004)	0.029 (0.002)
$\hat{\beta}$ **DNO Region: Scotland [Ref Midlands & South]**	—	0.015 (0.010)	0.017 (0.008)	0.023 (0.008)	0.008 (0.005)
Pre-treatment Covariates?	No	Yes	Yes	Yes	Yes
Observations	78,724	69,168	96,477	123,280	350,361
Estimator	OLS	OLS	OLS	OLS	OLS
Heteroscedasticity-Consistent Std. Errors (HC0)	Yes	Yes	Yes	Yes	Yes
$R^2_{\textbf{Adj.}}$	0.000	0.309	0.305	0.985	0.961

Note: The table presents parameter estimates and standard errors (parentheses) for the LATE ($\hat{\beta}$ Intraday Opt-in Notice [Ref Day-ahead Notice]), the expected average outcome in the control group ($\hat{\beta}$ Intercept), and partial correlations for pre-treatment covariates from models fit to subsets of our Saving Session data using asymmetric bandwidths h_{Left} and h_{Right} optimized to reduce mean-squared error (MSE) (see Section AI.7.2). Hour-specific *ATEs* ($\hat{\beta}$ Intraday Opt-in Notice (Time Range]) are from a model fit to a subset of our data obtained using only the left MSE-optimal bandwidth. Results rounded to three decimal places. See Long and Ervin (2000) for a discussion and comparison heteroscedasticity-consistent covariance matrices.

Table AT.27: Models of the probability of opting into the 10th Saving Session (regression discontinuity design).

MSE-Optimal Bandwidth ($h_{\text{Left}}, h_{\text{Right}}$)	($\mathbf{h_L}, \mathbf{h_R}$)	($\mathbf{h_L}, \mathbf{h_R}$)	($\mathbf{h_{L\div 1.5}}, \mathbf{h_{R\times 1.5}}$)	($\mathbf{h_{L\div 2}}, \mathbf{h_{R\times 2}}$)	$\mathbf{h_L}$ **Only**
$\hat{\beta}$ **Intercept**	0.563 (0.007)	0.566 (0.006)	0.566 (0.010)	0.572 (0.015)	0.576 (0.003)
$\hat{\beta}$ **Intraday Opt-in Notice**	0.002 (0.008)	-0.014 (0.008)	-0.019 (0.010)	-0.025 (0.015)	—
$\hat{\beta}$ **Intraday Opt-in Notice (07:59:59, 09:00:00]**	—	—	—	—	-0.026 (0.005)
$\hat{\beta}$ **Intraday Opt-in Notice (09:00:00, 10:00:00]**	—	—	—	—	-0.036 (0.005)
$\hat{\beta}$ **Intraday Opt-in Notice (10:00:00, 11:00:00]**	—	—	—	—	-0.045 (0.004)
$\hat{\beta}$ **Intraday Opt-in Notice (11:00:00, 12:00:00]**	—	—	—	—	-0.055 (0.004)
$\hat{\beta}$ **Intraday Opt-in Notice (12:00:00, 13:00:00]**	—	—	—	—	-0.053 (0.005)
$\hat{\beta}$ **Account ID (Millions) [Relative to Cutoff]**	-0.053 (0.021)	0.008 (0.028)	-0.019 (0.048)	0.031 (0.084)	—
$\hat{\beta}$ **Intraday Opt-in Notice** × **Account ID**	-0.027 (0.027)	-0.051 (0.026)	-0.009 (0.044)	-0.063 (0.082)	—
$\hat{\beta}$ **Avg. Hist. Sess. Consumption (kWh) [Z-Score]**	—	-0.127 (0.031)	-0.139 (0.029)	-0.119 (0.026)	-0.018 (0.010)
$\hat{\beta}$ **Avg. Hist. Sess. Participation [Z-Score]**	—	0.231 (0.001)	0.230 (0.001)	0.230 (0.001)	0.229 (0.001)
$\hat{\beta}$ **Total P376 Baseline (kWh) [Z-Score]**	—	0.089 (0.025)	0.100 (0.023)	0.084 (0.021)	0.014 (0.008)
$\hat{\beta}$ **Est. Annual Consumption (kWh) [Z-Score]**	—	-0.005 (0.004)	-0.005 (0.003)	-0.004 (0.003)	-0.002 (0.001)
$\hat{\beta}$ **IMD Rank (Postcode) [Z-Score]**	—	0.004 (0.002)	0.006 (0.001)	0.006 (0.001)	0.006 (0.001)
$\hat{\beta}$ **Tenure with Octopus (Years) [Z-Score]**	—	0.025 (0.019)	0.000 (0.021)	-0.000 (0.020)	-0.014 (0.003)
$\hat{\beta}$ **Business Entity [Ref Non-Business]**	—	0.072 (0.076)	0.028 (0.071)	0.043 (0.058)	0.009 (0.028)
$\hat{\beta}$ **Has Non-Octopus Product [Ref Octo. Prod.]**	—	0.241 (0.027)	0.232 (0.025)	0.234 (0.024)	0.235 (0.012)
$\hat{\beta}$ **Has Smart Tariff [Ref Non-Smart Tariff]**	—	0.030 (0.005)	0.039 (0.005)	0.042 (0.004)	0.030 (0.003)
$\hat{\beta}$ **DNO Region: North [Ref Midlands & South]**	—	-0.005 (0.003)	-0.003 (0.003)	-0.000 (0.003)	-0.001 (0.002)
$\hat{\beta}$ **DNO Region: Scotland [Ref Midlands & South]**	—	0.000 (0.006)	0.003 (0.006)	0.005 (0.005)	0.005 (0.003)
Pre-treatment Covariates?	No	Yes	Yes	Yes	Yes
Observations	99,678	88,422	104,422	125,131	377,569
Estimator	OLS	OLS	OLS	OLS	OLS
Heteroscedasticity-Consistent Std. Errors (HC0)	Yes	Yes	Yes	Yes	Yes
$R^2_{\text{Adj.}}$	0.001	0.224	0.220	0.219	0.215

Note: The table presents parameter estimates and standard errors (parentheses) for the LATE ($\hat{\beta}$ Intraday Opt-in Notice [Ref Day-ahead Notice]), the expected average outcome in the control group ($\hat{\beta}$ Intercept), and, partial correlations for pre-treatment covariates from models fit to subsets of our Saving Session data using asymmetric bandwidths h_{Left} and h_{Right} optimized to reduce mean-squared error (MSE) (see Section AI.7.2). Hour-specific *ATEs* ($\hat{\beta}$ Intraday Opt-in Notice (Time Range]) are from a model fit to a subset of our data obtained using only the left MSE-optimal bandwidth. Results rounded to three decimal places. See Long and Ervin (2000) for a discussion and comparison heteroscedasticity-consistent covariance matrices.

AT.8 Field Trial Regression Output: Complete Results

Table AT.28: Models of total consumption (kWh) during the 12th Saving Session (field trial).

Equation Response Variable	Reduced Fm. Consumption	Reduced Fm. Consumption	Reduced Fm. Consumption	1^{st} Stage Intraday SMS + £	2^{nd} Stage Consumption	2^{nd} Stage Consumption	2^{nd} Stage Consumption
$\hat{\beta}$ **Intercept**	0.655 (0.001)	0.656 (0.001)	0.662 (0.001)	-0.000 (0.000)	0.655 (0.001)	0.656 (0.001)	0.662 (0.001)
$\hat{\beta}$ **Day-ahead Email**	-0.021 (0.006)	-0.018 (0.006)	-0.011 (0.005)	0.000 (0.000)	-0.021 (0.006)	-0.018 (0.006)	-0.011 (0.005)
$\hat{\beta}$ **Intraday SMS + £1.25 Incentive Assigned**	-0.007 (0.006)	-0.007 (0.006)	-0.007 (0.005)	0.228 (0.003)	—	—	—
$\hat{\beta}$ **Intraday SMS + £1.25 Incentive**	—	—	—	—	-0.030 (0.025)	-0.033 (0.026)	-0.029 (0.021)
$\hat{\beta}$ **Avg. Hist. Sess. Consumption (kWh) [Z-Score]**	—	—	1.017 (0.020)	0.002 (0.001)	—	—	1.017 (0.020)
$\hat{\beta}$ **Avg. Hist. Sess. Participation [Z-Score]**	—	—	-0.021 (0.001)	-0.000 (0.000)	—	—	-0.021 (0.001)
$\hat{\beta}$ **Total P376 Baseline (kWh) [Z-Score]**	—	—	1.179 (0.012)	-0.001 (0.000)	—	—	1.179 (0.012)
$\hat{\beta}$ **Est. Annual Consumption (kWh) [Z-Score]**	—	—	0.058 (0.004)	-0.000 (0.000)	—	—	0.058 (0.004)
$\hat{\beta}$ **IMD Rank (Postcode) [Z-Score]**	—	—	-0.002 (0.001)	-0.000 (0.000)	—	—	-0.002 (0.001)
$\hat{\beta}$ **Tenure with Octopus (Years) [Z-Score]**	—	—	-0.010 (0.001)	0.001 (0.000)	—	—	-0.009 (0.001)
$\hat{\beta}$ **Business Entity [Ref Non-Business]**	—	—	-0.010 (0.072)	0.002 (0.004)	—	—	-0.009 (0.072)
$\hat{\beta}$ **Has Non-Octopus Product [Ref Octo. Prod.]**	—	—	-0.041 (0.004)	-0.005 (0.000)	—	—	-0.041 (0.004)
$\hat{\beta}$ **Has Smart Tariff [Ref Non-Smart Tariff]**	—	—	-0.040 (0.004)	0.001 (0.000)	—	—	-0.040 (0.004)
$\hat{\beta}$ **DNO Region: North [Ref Midlands & South]**	—	—	0.012 (0.002)	0.000 (0.000)	—	—	0.012 (0.002)
$\hat{\beta}$ **DNO Region: Scotland [Ref Midlands & South]**	—	—	0.012 (0.004)	0.001 (0.000)	—	—	0.012 (0.004)
Pre-treatment Covariates?	No	No	Yes	Yes	No	No	Yes
Observations	638,242	540,395	540,395	540,395	638,242	540,395	540,395
Estimator	OLS	OLS	OLS	OLS	IV-2SLS	IV-2SLS	IV-2SLS
Heteroscedasticity-Consistent Std. Errors (HC0)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$R^2_{\text{Adj.}}$	0.000	0.000	0.367	0.223	0.000	0.000	0.367
Partial F-Statistic				4728.325			
Wooldridge's Exogeneity Test p-value					0.433	0.554	0.485

Note: The table presents parameter estimates and standard errors (parentheses) for the ATE ($\hat{\beta}$ Intraday Notice + Day-ahead Email [Ref Intraday Only]), the CACE ($\hat{\beta}$ Intraday Notice + Intraday SMS + £1.25 Incentive [Ref Intraday Only]), the ITT ($\hat{\beta}$ Intraday SMS + £1.25 Incentive Assigned), the expected average outcome in the control group ($\hat{\beta}$ Intercept), and partial correlations for pre-treatment covariates from reduced form ordinary least-squares (OLS) regression models as well as the 1^{st} and 2^{nd} stages of two-stage-least-squares (2SLS) regression models. "Intraday SMS + £1.25 Incentive Assigned" is the singular instrumental variable (IV). Results rounded to three decimal places. See Long and Ervin (2000) for a discussion and comparison heteroscedasticity-consistent covariance matrices. H_0 for Wooldridge's regression text of exogeneity is that the endogenous variable "Intraday SMS + £1.25 Incentive Assigned" is exogenous. See Table AT.25 for descriptive statistics and reference categories.

Table AT.29: Models of the probability of opting into the 12th Saving Session (field trial).

Equation Response Variable	Reduced Fm. Participation	Reduced Fm. Participation	Reduced Fm. Participation	1^{st} Stage Intraday SMS + £	2^{nd} Stage Participation	2^{nd} Stage Participation	2^{nd} Stage Participation
$\hat{\beta}$ **Intercept**	0.443 (0.001)	0.434 (0.001)	0.422 (0.001)	-0.000 (0.000)	0.443 (0.001)	0.434 (0.001)	0.422 (0.001)
$\hat{\beta}$ **Day-ahead Email**	0.026 (0.004)	0.026 (0.004)	0.025 (0.004)	0.000 (0.000)	0.026 (0.004)	0.026 (0.004)	0.025 (0.004)
$\hat{\beta}$ **Intraday SMS + £1.25 Incentive Assigned**	0.024 (0.004)	0.022 (0.004)	0.022 (0.004)	0.227 (0.003)	—	—	—
$\hat{\beta}$ **Intraday SMS + £1.25 Incentive**	—	—	—	—	0.103 (0.016)	0.095 (0.017)	0.098 (0.016)
$\hat{\beta}$ **Avg. Hist. Sess. Consumption (kWh) [Z-Score]**	—	—	0.014 (0.003)	0.001 (0.000)	—	—	0.014 (0.003)
$\hat{\beta}$ **Avg. Hist. Sess. Participation [Z-Score]**	—	—	0.202 (0.001)	-0.000 (0.000)	—	—	0.202 (0.001)
$\hat{\beta}$ **Total P376 Baseline (kWh) [Z-Score]**	—	—	-0.013 (0.003)	-0.001 (0.000)	—	—	-0.013 (0.003)
$\hat{\beta}$ **Est. Annual Consumption (kWh) [Z-Score]**	—	—	-0.007 (0.001)	-0.000 (0.000)	—	—	-0.007 (0.001)
$\hat{\beta}$ **IMD Rank (Postcode) [Z-Score]**	—	—	-0.002 (0.001)	-0.000 (0.000)	—	—	-0.002 (0.001)
$\hat{\beta}$ **Tenure with Octopus (Years) [Z-Score]**	—	—	0.041 (0.001)	0.001 (0.000)	—	—	0.041 (0.001)
$\hat{\beta}$ **Business Entity [Ref Non-Business]**	—	—	0.107 (0.025)	0.002 (0.004)	—	—	0.107 (0.025)
$\hat{\beta}$ **Has Non-Octopus Product [Ref Octo. Prod.]**	—	—	0.108 (0.003)	-0.005 (0.000)	—	—	0.109 (0.003)
$\hat{\beta}$ **Has Smart Tariff [Ref Non-Smart Tariff]**	—	—	0.073 (0.002)	0.001 (0.000)	—	—	0.073 (0.002)
$\hat{\beta}$ **DNO Region: North [Ref Midlands & South]**	—	—	-0.009 (0.001)	0.000 (0.000)	—	—	-0.009 (0.001)
$\hat{\beta}$ **DNO Region: Scotland [Ref Midlands & South]**	—	—	0.008 (0.003)	0.001 (0.000)	—	—	0.008 (0.003)
Pre-treatment Covariates?	No	No	Yes	Yes	No	No	Yes
Observations	650,809	551,494	551,494	551,494	650,809	551,494	551,494
Estimator	OLS	OLS	OLS	OLS	IV-2SLS	IV-2SLS	IV-2SLS
Heteroscedasticity-Consistent Std. Errors (HC0)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$R^2_{\text{Adj.}}$	0.000	0.000	0.174	0.223	0.001	0.001	0.175
Partial F-Statistic				4814.807			
Wooldridge's Exogeneity Test p-value					0.021	0.021	0.010

Note: The table presents parameter estimates and standard errors (parentheses) for the ATE ($\hat{\beta}$ Intraday Notice + Day-ahead Email [Ref Intraday Only]), the CACE ($\hat{\beta}$ Intraday Notice + Intraday SMS + £1.25 Incentive [Ref Intraday Only]), the ITT ($\hat{\beta}$ Intraday SMS + £1.25 Incentive Assigned), the expected average outcome in the control group ($\hat{\beta}$ Intercept), and partial correlations for pre-treatment covariates from reduced form ordinary least-squares (OLS) regression models as well as the 1^{st} and 2^{nd} stages of two-stage-least-squares (2SLS) regression models. "Intraday SMS + £1.25 Incentive Assigned" is the singular instrumental variable (IV). Results rounded to three decimal places. See Long and Ervin (2000) for a discussion and comparison heteroscedasticity-consistent covariance matrices. H_0 for Wooldridge's regression text of exogeneity is that the endogenous variable "Intraday SMS + £1.25 Incentive Assigned" is exogenous. See Table AT.25 for descriptive statistics and reference categories.

AT.9 Cost Effectiveness: Comparison Between P376 and DiD Reductions per Saving Session

Table AT.30: Cost to NGESO per MWh of DFS utilization using clipped P376 methodology, our DiD methodology and marginal balancing mechanism actions.

Session	Total	1	2	3	4	5	6	7	8	9	10	11	12	13
Official demand reduction of all DFS providers (MWh)	3157.61	133.39	192.81	216.92	211.61	398.67	1771.19	396.8	283.17	175.48	241.55	269	231.36	235.7
Official demand reduction non Octopus Energy DFS providers (using clipped P376) (MWh)	1294.4	23.88	82.62	101.55	103.67	120.79	89.3	209.23	35.26	76.45	105	115.13	120.48	111.06
Octopus Energy official demand reduction (using clipped P376) (MWh)	1863.21	109.51	110.19	115.37	107.94	277.88	81.89	187.56	247.91	99.03	136.55	153.87	110.88	124.64
DiD estimate of "actual" demand reduction by Octopus Energy (MWh)	1642.09	93.02	121.69	111.46	92.83	277.57	68.80	185.99	216.33	67.89	107.45	106.94	113.67	78.42
Costs of DFS procurement of Octopus Energy customer demand reduction (k£)[a]	6186.36	328.55	330.56	346.1	323.81	833.63	245.67	912.79	990.38	297.08	409.65	461.6	332.64	373.91
Cost per MWh "actual" demand reduction (£/MWh)	4580.09	3482.47	3318.6	3838.85	4026.56	4532.78	4488.35	5774.05	5159.59	4425.74	4321.21	4712.1	4649.88	5658.61
Cost of marginal Balancing Mechanism action (£/MWh)[b]	1152.52	335	980.89	695	590	5500	350.63	412.5	512.67	277	291	478	253.25	179.5

Note: We show a series of connected facts in this table. First, Octopus Energy contributed 59% of the total demand reduction of the Demand Flexibility Service on the events that Octopus Energy participated in (where each DFS provider measured its reduction using the NGESO's prescribed methodology). Second, comparing our DiD methodology to NGESO's ("clipped" pre-post-style) methodology – only for Octopus Energy customers, during the 13 DFS events Octopus energy participated in – suggests that the official methodology typically overestimates demand reduction. However, this pattern is not universally true; in the second and twelfth sessions, our demand reduction estimates are larger than the official results. Overall, our DiD methodology shows a demand reduction of 1642 MWh from Octopus Energy customers, lower than the official demand reduction of 1863 MWh reported by the P376 methodology. Third, when we compare the costs to NGESO given our estimate of "actual" demand reduction, we found slightly higher costs than the official £3,000/MWh that ESO procured under the assumption of demand reduction according to the official methodology. Fourth, the marginal Balancing Mechanism action cost is substantially lower than the cost of demand reduction the DFS provided during the 29 half-hours in which Saving Sessions took place, except for the fifth Saving Session on December 12, 2022, when the marginal balancing action's cost reached £5,500/MWh.

[a] We use Settled Volume (MWh) and expenditure (£) reported by NGESO in the National Grid Data Portal for DFS Test and Live events. The settled costs of Octopus Energy customer demand reduction is calculated by multiplying the total settled costs by the proportion of Octopus Energy delivery over total delivery (National Grid, 2023*g*,*h*).

[b] Prices are from LCP Delta (2023) and calculated as the average marginal price over the settlement periods in which the events took place.

AT.10 MVPF: Assumed NGESO Costs Based on Saving Sessions Payments to Customers

Table AT.31: Rewards given to Octopus Energy customer for each type of event.

Type of event	OctoPoints awarded per kWh	Value [£] (1 OP =0.00125£/kWh)	NGESO payments (£/kWh)
Test event	1,800	2.25	3
Live event - Jan 23, 2023	2,700	3.375	4.125
Live event - Jan 24, 2023	3,200	4	4.75

Note: NGESO provided a Guaranteed Acceptance Price of £3 for every kWh of demand reduction during Test events calculated with the P376 methodology to all the retailers which participated in the Demand Flexibility Service. All retailers then decided how to incentivise their customers for their demand reduction with various incentive structures. Octopus Energy rewarded customer with OctoPoints, a currency that could be used directly to transfer funds into their Octopus Energy account or be converted directly into cash. For each kWh of demand reduction, Octopus Energy customers received 1,800 OctoPoints, equivalent to 2.25 £/kWh for all the Test events. For Live events, prices were calculated through a private auction mechanism, so these prices are not publicly available for each supplier. OE awarded 2,700 and 3,200 OctoPoints for the Live events held on 2023-01-23 and 2023-01-24 respectively. We assume that Octopus Energy retained the same amount of money per customer per kWh of demand reduction – £0.75. We thus obtain assumed auction prices of £4.125 and £4.75.

AT.11 MVPF: Greenhouse Gas Emission Factors and Damage Costs

Table AT.32: GHG emissions factors and damage costs for NO_x, SO_x and PM2.5 pollutants.

Pollutant emitted	Gas emission factor [tonne/MWh]	Coal emission factor [tonne/MWh]	Damage costs [$£_{2023}$/tonne]
$\mathbf{NO_x}$	$9.00e-05$	$3.24e-04$	8,148
$\mathbf{SO_x}$	$3.60e-06$	$2.92e-04$	16,616
PM2.5	$2.20e-07$	$2.23e-06$	74,769

Note: Emission factors represent the amount of pollutants released in a particular activity. In this case, we use the emissions factors for NO_x, SO_x and PM2.5 from Department for Environment, Food and Rural Affairs (Defra) (2023) related to the production of electricity using natural gas or coal as the main fuel. To asses the the air quality impact of our campaign, we need to use the damage costs of each of these pollutants from UK GOV (*2023a*). These represent a set of monetary impact values per tonne of emission.

AT.12 MVPF: Comparison of MVPFs Assuming Different Demand Reduction Methodologies

Table AT.33: Values of MVPF for each individual Saving Session and for the whole program.

Day	MVPF - DiD	MVPF - P376	MVPF (VoLL scenario) - DiD	MVPF (VoLL scenario) - P376
November 15, 2022	1.06	1.08	2.76	3.08
November 22, 2022	1.07	1.08	3.29	3.08
November 30, 2022	1.06	1.08	3.01	3.08
December 1, 2022	1.06	1.08	2.78	3.08
December 12, 2022	1.02	1.04	3.03	3.04
January 19, 2023	1.05	1.08	2.74	3.08
January 23, 2023	1.02	1.03	2.47	2.48
January 24, 2023	1.02	1.02	2.12	2.29
January 2023	1.05	1.08	2.42	3.08
February 13, 2023	1.05	1.08	2.63	3.08
February 21, 2023	1.05	1.08	2.44	3.08
March 15, 2023	1.05	1.08	3.13	3.08
March 23, 2023	1.04	1.05	2.31	3.08
Total	**1.05**	**1.05**	**2.63**	**2.85**

Note: The MVPF of Saving Sessions we calculate using the costs and benefits outlined in Sections 5.3.1 and 5.3.2. In columns 2 and 4, we use the demand reduction derived from our Octopus versus Bulb DiD to value the CO_2 carbon emissions abatement and value of lost load. In columns 3 and 5, we show the MVPF where demand reduction is assumed to be the estimate from NGESO's official methodology ("P376").