Who heeds the call to conserve in an energy emergency? Evidence from smart thermostat data

Dylan Brewer & Jim Crozier*

February 3, 2022

Abstract

In 2019, a fire at a natural gas plant and historically low temperatures caused an emergency shortage of natural gas in Michigan. To avoid an outage, the Governor issued a request via statewide text alert to turn thermostats down to 65°F. We analyze the effectiveness of this request using high frequency smart-thermostat data from Michigan and four neighboring states. Using a difference-in-differences research design, we find that Michigan households reduced thermostat settings by 0.86 degrees on average following the governor's request. Households that were previously above 65°F responded strongly, while households that were below did not respond at all. Meanwhile, households in districts that voted for the Governor in 2018 were more likely to comply. Our results suggest that unrealistic compliance goals and political polarization reduce the effectiveness of emergency calls to conserve energy.

Keywords: behavioral economics, nudges, moral suasion, energy use, natural gas, natural disasters, reference points, natural field experiments

*School of Economics, Georgia Institute of Technology, 221 Bobby Dodd Way, Atlanta, Georgia 30332, brewer@gatech.edu and rcrozier3@gatech.edu. We gratefully acknowledge the support and contribution of Ecobee and Ecobee customers to this research. This material is based upon work supported by the Google Cloud Research Credits program with the award GCP19980904. We thank Soren Anderson, Laura Taylor, A. Justin Kirkpatrick, Prabhat Barnwal, Alecia Cassidy, Casey Wichman, Nathan Chan, Elise Breshears, Matt Oliver, Cody Orr, and seminar participants at Virginia Tech, the Southern Economics Association Meetings, and the UCLA Climate Adaptation Research Symposium for useful discussion and feedback. Thank you to Graham Lewis for research assistance. Employees at Consumers Energy provided valuable perspective and total gas consumption data, for which we are grateful.

1 Introduction

During emergencies, government officials often make appeals to citizens to contribute effort toward a common goal. These appeals have taken the form of requests to contribute effort toward a public good (such as buying war bonds), to voluntarily ration consumption of scarce goods (such as reducing consumption of water during a drought), or to comply with safety protocols (such as taking certain precautions during a pandemic). Requests have become sophisticated over time as the communication medium has evolved from print materials, to radio and television announcements, and now to digital alerts with the ability to target specific individuals in real time. The strategy of requests has also evolved with advances in psychological research in how to influence behavior (Cialdini, 2006) and the use of the "nudge" framework to reduce the costs of compliance (Thaler and Sunstein, 2008). Modern requests for pro-social behavior often take the form of nudge reminders, informational treatments (Allcott and Taubinsky, 2015), appeals to morality or "moral suasion" (Ito et al., 2018), appeals to expert authority (Breza et al., 2021), or social comparison to peer behavior to induce compliance (Ferraro et al., 2011; Allcott and Rogers, 2014).

This paper analyzes the efficacy of an emergency request by the Michigan Governor for households to reduce thermostat settings during a natural gas shortage caused by a fire at a natural gas facility. During the cold wave of the 2019 polar vortex, outdoor temperatures were extremely low, causing high demand for natural gas for space heating. At 10:30 am on January 30, 2019, a fire broke out at Consumers Energy's largest natural gas storage facility. Consumers Energy is a gas and electric utility that serves roughly half of Michigan's households. By 1:00 pm, the utility had recognized that demand for natural gas may exceed supply, with the potential to cause the system to fail. At 2:30 pm, the utility requested via emails, social media, and news media that all households reduce natural gas consumption. At 10:00 pm as pipeline pressures continued to drop, the Michigan Governor tweeted a request to conserve natural gas and at 10:30 pm followed up with an emergency alert directly to cell phones in Michigan requesting that households reduce thermostat settings to 65°F or below. The utility communications ensured that at least some households were aware of the request, but the cell phone alert went out to all households within the lower peninsula of Michigan. The next day at 4:30 pm, the utility issued an “all clear” time of midnight to its customers—thanks to voluntary reductions in demand by households and industrial consumers, the system did not fail and natural gas outages were avoided.

To measure household responses to the requests, we use smart thermostat data provided by Ecobee’s Donate-Your-Data program. The data include thermostat setting and furnace fan run time at 5-minute intervals. This high-frequency, household-level data allows us to observe and measure each household’s response to the requests as the emergency unfolded. Our empirical strategy uses households in the surrounding states of Ohio, Indiana, Illinois, and Wisconsin as control units for a difference-in-differences approach. We find that mean thermostat settings, the proportion of homes with a thermostat setting below 65 degrees Fahrenheit, and furnace fan run time exhibit parallel trends across the treatment and control units, which supports our interpretation of our estimates as a causal effect of the emergency request on changes in thermostat settings.

Using a difference-in-differences strategy with four control states, we estimate the average treatment effect of the emergency request. We find that on average, households lowered their thermostats by 0.86 degrees Fahrenheit, roughly 20% of the size of the typical variation in the average thermostat setting. The request increased the proportion of household thermostat settings below 65°F by 11 percentage points, a 46% increase relative to the proportion of households whose thermostat settings are normally below 65°F. Finally, we examine the effect of the request on furnace fan run time, which is our best available proxy for household natural gas consumption. We find evidence that the emergency request reduced the furnace run time by 2 minutes per hour, an 8.2% reduction relative to the conditional mean run time for Michigan households during the emergency. These results are robust across a number of specifications and checks for spillover treatment to border counties, and a placebo test with an earlier cold wave shows that this behavior is not driven by outside temperature alone. An event-study analysis reveals that the Governor’s amplification of the utility’s earlier request nearly quadrupled compliance rates. Prior to the Governor’s alert, only 5%

of additional households reduced thermostat settings below 65°F; after the Governor's alert, the additional compliance rate was as high as 18%. We interpret this as evidence that the Governor's authority was essential to increasing the salience of the appeal.

The utility and Governor framed the emergency request with a clear reference point of 65°F that affected household responses. This creates two types of reference point heterogeneity: households that normally set the thermostat below this point were essentially exempted from the emergency request, and households that normally set the thermostat significantly higher were asked to deviate more from their typical consumption patterns. We observe strong perverse framing effects in the data (Tversky and Kahneman, 1981). Using a nonparametric estimate of each household's expected baseline temperature, we find that households that typically set the thermostat below 65°F were unresponsive to the emergency request on average. Households that are typically the coldest (far below 65°F) increased the thermostat after the emergency request. The higher a household's baseline thermostat setting above the reference point, the less likely the household was to comply with the request. The results suggest that setting a more aggressive reference point trades off an increased treatment effect for individuals near the reference point with decreased compliance from discouraged individuals far from the reference point.

We scrutinize the role of political polarization as a factor in determining compliance with the request. The Governor, Gretchen Whitmer, assumed office in January 2019, less than a month before the polar vortex. We hypothesize that households that did not approve of the Governor may have been less likely to comply with the request. Our analysis studies the differential compliance of households in counties that supported the Governor's 2018 election bid, using data on county-level election returns. We show that compliance rates and the average reduction in thermostat setting are increasing in the Governor's vote share. Households in counties where the Governor's vote share was above 70% were three times more likely to comply with the request than households in counties where the Governor's vote share was below 40%.

The results of this study are important for policymakers studying compliance with emergency requests in a broad range of fields. For instance, during the COVID-19 pandemic, local, national, and international governmental bodies sought to coordinate behavior to reduce the spread of the virus through a combination of compulsory policies and requests for voluntary compliance. Pandemic-related policies and requests were met with mixed compliance and even open defiance, and a growing literature seeks to understand the effects of political affiliation on cooperation with requests for social distancing and stay-at-home orders (e.g., Allcott et al., 2020; Barrios and Hochberg, 2020). Related to compliance with energy and environmental policy, economists have studied firm strategic avoidance of air quality monitoring (Zou, 2021), imperfect enforcement of emissions caps (Sigman and Chang, 2011), voluntary reductions of emissions (Foster et al., 2009; Foster and Gutierrez, 2013), compliance with the US acid rain program (Montero, 1999), and the use of regulatory loopholes to avoid compliance with fuel efficiency regulations (Anderson and Sallee, 2011). Beatty et al. (2019) study household emergency preparedness for hurricanes, finding that household behavior is highly influenced by recent hurricane events and that households in general do not follow government preparedness recommendations. Other work shows that an increased perception of risk and confidence in government institutions increases compliance with hurricane evacuation orders (Whitehead et al., 2000; Kim and Oh, 2015). In another context, Wichman et al. (2016) find that households in North Carolina reduced consumption of water during a drought when both voluntary and mandatory non-price mechanisms were implemented to restrict water use. Our paper contributes to these literatures by providing what we believe is the most granular data on household compliance with emergency requests. In addition, the unexpected nature of the emergency and its isolation to one state creates a credible natural experiment that allows us to pursue an identification strategy that takes advantage of plausibly exogenous time and cross-sectional variation, which is uncommon for this literature.

Our work contributes to the empirical literature studying reference points and economic behavior. Research in this area examines labor supply behavior relative to earnings expectations (Thakral and Tô, 2021; Farber, 2008; Camerer et al., 1997), retirement decisions relative to age reference points (Seibold, 2021), and loss aversion in tax filing (Engström et al., 2015). Other work focuses on the use of social comparison as a reference point to influence behavior (e.g., Allcott (2011), Ferraro and Price (2013), Brent et al. (2015), and Hallsworth et al. (2017)). In the charitable giving literature, suggested donation amounts increase voluntary contributions and anchor donations to the suggested amount (Edwards and List, 2014). Harding and Hsiaw (2014) study how non-binding goal setting for energy conservation leads to behavior consistent with reference-dependent preferences. Brown et al. (2013) find that factory-default thermostat settings substantially impacted subsequent thermostat levels chosen in the workplace. Our paper contributes to this literature by studying a novel reference point created within the phrasing of a governmental emergency request. Our results suggest that for the policymaker, setting a reference point more aggressively trades off an increase in the effect of meeting the reference point with the cost of meeting the reference point. In our context, further lowering the requested thermostat setting would have reduced compliance from those with high baseline settings but would have increased the effort from those with medium and low baseline settings. These findings imply that policies may be designed so as to have effect-maximizing reference point levels.

In addition, this paper is relevant to the literature in environmental and energy economics analyzing the use of non-price mechanisms to conserve household consumption of water, natural gas, and electricity.[1] Given political constraints on raising prices of these goods, regulators and suppliers have sought to curb consumption via mandatory restrictions and voluntary requests. In Ito et al. (2018), the authors conduct a field experiment that provided Japanese households with voluntary appeals or price incentives to reduce electricity consumption. Relative to a control group, voluntary appeals resulted in a short term reduction in electricity consumption of 8% while price incentives resulted in a reduction in electricity consumption of 17% that was sustained over a longer period. We find a similar effect size for the voluntary appeal in our emergency setting. Allcott (2011), Ferraro et al. (2011), Ferraro and Price (2013) and Brent et al. (2015) study the use of social comparisons,

[1] See Carlsson et al. (2021) for a recent overview of papers analyzing nudges and non-price mechanisms.

finding that households voluntarily reduce electricity and water consumption when told how much similar households are consuming. Allcott and Rogers (2014) find evidence that households reduce electricity consumption in response to home energy reports and that repeated treatments induce additional reductions and enforce habits. Kotchen et al. (2012) find that high-consumption residential electricity consumers are more likely to voluntarily contribute to a public good by purchasing "green" electricity at a higher rate. Luyben (1982) studies a 1977 request by President Carter for US households to reduce thermostat settings to 65°F or below, finding that compliance was low overall (27%) and that self-reported compliance was higher than recorded compliance. Our findings provide new evidence that political context substantially impacts the effectiveness of non-price mechanisms. We also provide novel estimates of the effect of more costly non-price mechanisms on compliance that suggests households trade off the cost of compliance with a moral cost in a manner consistent with the framework in Levitt and List (2007).

Our paper proceeds by describing the polar vortex and natural gas fire events in greater detail. We then introduce the smart thermostat data used in the paper. Next, we present our empirical strategy and analysis, which we subdivide into a section estimating the average treatment effect of the request, a section presenting an event-study analysis, a section examining the effects of the reference point on behavior, and a section examining the role of political support of the Governor on compliance. Finally, section 5 summarizes the findings and concludes.

2 Polar vortex and natural gas fire events

Extreme cold weather events caused by disturbances to the polar vortex have recently received significant attention in the United States and Europe. Perhaps most notably was the 2021 polar vortex event that overwhelmed the electricity grid in Texas, killing 172 people and resulting in damages valued at levels ranging from $20 billion to $295 billion (NOAA, 2021; Perryman Group, 2021). Since 1980, winter disasters have resulted in 19 "billion-dollar climate disasters" in the United States, causing 1,223 deaths (NOAA, 2021). There is only weak evidence that climate change is contributing to the perceived increase in polar vortex events (Blackport and Screen, 2020); however, aging energy infrastructure in the United States and Europe may increase the costs of such events in the future.

Beginning on Tuesday, January 29, 2019, temperatures in the Midwest declined to nearly record-low levels as cold air in the stratosphere over the Arctic blew southward over North America (NOAA, 2019). Temperatures reached -23°F in Chicago, -13°F in Detroit, -11°F in Indianapolis, and as low as -45°F elsewhere in the United States (EIA, 2019b). On Wednesday, January 30, 2019, single-day estimated natural gas consumption in the United States hit an all-time high with 37.9 billion cubic feet consumed in a single day, and electricity demand in the Midwest approached all-time peaks (EIA, 2019a). In 2017, over 75 percent of Michigan homes used natural gas as the primary heating fuel (MPSC, 2019b).

Coinciding with this extreme demand-side stress, a supply-side emergency caused a near system-wide natural gas delivery failure in Michigan. On January 30, 2019 at 10:30 am, a fire broke out at the Ray Compressor Station in Macomb County, Consumers Energy's largest natural gas storage facility (MPSC, 2019b). Immediately after the fire broke out, the utility drew upon standby natural gas reserves to stabilize pipeline pressures (Consumers Energy Company, 2019). By 1:00 pm, Consumers Energy recognized the possibility that demand could exceed supply, which could cause total system failure, and contacted their highest demand industrial and commercial customers with requests to reduce consumption of natural gas. At 2:26 pm, Consumers issued a tweet requesting households to reduce thermostat settings and sent emails to residential and business customers requesting reductions in natural gas use. Shortly thereafter, the CEO of Consumers Energy took to Facebook Live to urge households to reduce thermostat settings.[2] The utility ultimately sent over 500,000 external emails, made 21 social media posts, and responded to 130 media inquiries on January 30-31 (Consumers Energy Company, 2019). State-operated buildings reduced thermostat settings by 5°F and manufacturers reduced consumption of natural gas (Des-

[2]Tweet available at https://twitter.com/ConsumersEnergy/status/1090692811081551885 and Facebook Live video available at https://www.facebook.com/85543026043/videos/357785638397997/.

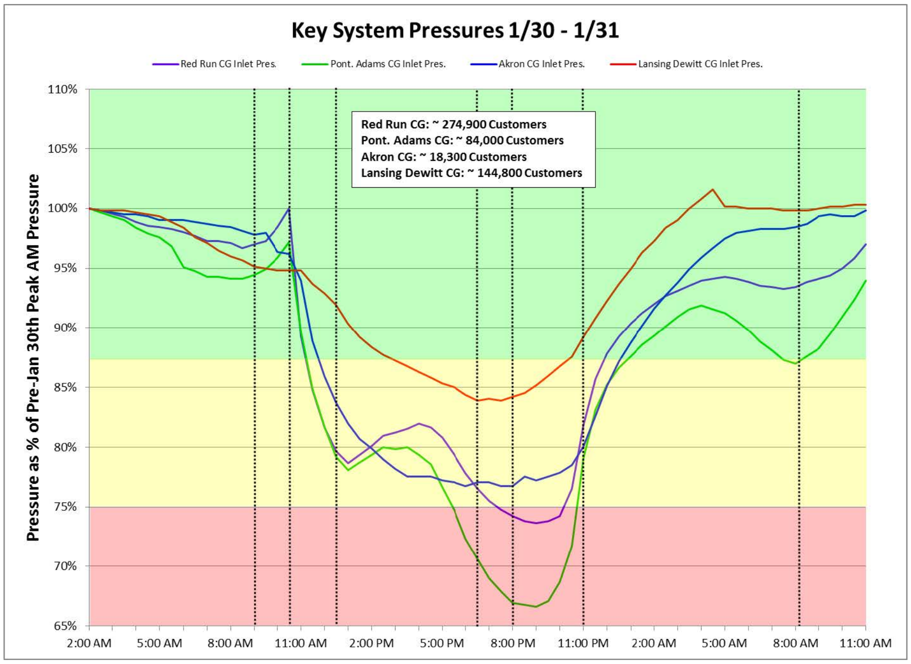


Figure 1: Each series corresponds to a Michigan natural gas pipeline instantaneous pressure, January 30-31, 2019. Image source: Michigan Public Service Commission Case No. U-20463 (Consumers Energy Company, 2019).

Ormeau, 2019). In addition, the utility issued mandatory curtailment orders for industrial and commercial natural gas customers and requested that natural gas electricity generators reduce generation to preserve residential heating (Consumers Energy Company, 2019). On the supply side, Consumers Energy purchased 925 MMcf/day worth of same-day supply of natural gas for January 30th, of which only 61% was ultimately delivered due to supply constraints (Consumers Energy Company, 2019).[3]

Despite efforts to reduce non-residential consumption and to procure natural gas on the supply side, the system was still unpacking (losing pressure) going into the evening. Figure 1 displays Michigan natural gas pipeline pressures on January 30th and 31st. Despite efforts to curb demand and increase supply, equilibrium pressures were dropping as the evening approached and temperatures continued to get colder. At 8:00 pm, Consumers Energy

[3]Same-day natural gas delivery is relatively rare compared to same-day electricity generation, for example. This event was the first time that Consumers Energy had attempted to secure same-day delivery (Consumers Energy Company, 2019). For extreme-weather events, utilities can increase pressure in natural gas pipelines ahead of time, storing gas within the system. Given that the flow of gas is not instantaneous, same-day supply is not typically used to balance supply and demand.

reached out to Governor Whitmer to make a final public appeal to households to reduce thermostat settings (Consumers Energy Company, 2019).

At 10:01 pm, the Governor of Michigan tweeted a request for households to reduce thermostats to 65°F, and at 10:30 pm activated FEMA's Wireless Emergency Alert system to send a text alert to all cell phones in Michigan asking households to reduce thermostat settings to 65°F (Gray, 2019).[4] The text of the cell phone alert message read "Due to extreme temps Consumers asks everyone to lower their heat to 65 or less through Fri." Conversations with the Michigan State Police Emergency Management and Homeland Security department and Consumers Energy indicated that officials believed 65°F was achievable, comfortable, and likely to be lower than the usual thermostat setting, but the number was chosen arbitrarily.

Shortly after the Governor's text message at 10:40 pm, 30% of the Ray Compressor Station capacity came back online, which combined with demand reductions to begin to increase pressures (Consumers Energy Company, 2019). Using data provided by Consumers Energy, forecasted natural gas demand using realized weather conditions was 3.3 billion cubic feet on January 30th and 2.9 billion cubic feet on January 31st. After all reductions in consumption were accounted for, actual consumption was 3.0 billion cubic feet on January 30th and 2.6 billion cubic feet on January 31st, implying a 10.7% and 10.5% reduction in daily consumption from all sources (residential and non-residential). On January 31st at 4:30 pm, Consumers Energy tweeted an "all clear" time of midnight that night, after which households could resume heating normally.[5]

Did households listen and comply with the emergency requests issued by the utility and public officials? Furthermore, how did the phrasing of the request around a thermostat setting of 65°F affect household compliance? Given the Governor's request, did political polarization affect which households were likely to comply? We answer these questions using high-frequency smart thermostat data.

[4]Tweet available at https://twitter.com/GovWhitmer/status/1090807363811065857.
[5]Tweet available at https://twitter.com/ConsumersEnergy/status/1091086892592959495.

Table 1: Summary statistics for households in Michigan and in the control states (Ohio, Indiana, Illinois, and Wisconsin).

	(1) Michigan mean/sd	(2) Controls mean/sd	(3) Difference b/t
Sq ft	2,291.68	2,442.15	150.47**
	(946.50)	(1,041.26)	(197.90)
Unit detached	0.54	0.52	-0.02**
	(0.50)	(0.50)	(-49.95)
Age of home (years)	34.25	34.93	0.68**
	(29.12)	(30.89)	(30.57)
Number of occupants	1.23	1.41	0.19**
	(1.71)	(1.76)	(144.00)
January 1 - 29 thermostat setting	66.88	67.37	0.50**
	(3.89)	(3.61)	(149.81)
January 30 thermostat setting before event	67.43	67.80	0.37**
	(3.76)	(3.80)	(16.18)
January 30 - 31 thermostat setting during event	66.23	67.54	1.32**
	(3.80)	(3.84)	(89.11)
January 1 - 29 outside temperature	24.24	25.38	1.14**
	(11.21)	(12.88)	(113.49)
January 30 - 31 outside temperature	-0.25	-0.43	-0.18**
	(1.66)	(3.03)	(-115.28)
N	2,372,193	7,053,172	9,425,365
Households	2,700	7,953	10,654

** p<0.01, * p<0.05

3 Data and research design

We use data on smart thermostat temperature settings provided by Ecobee as part of the 2019 release of the "Donate Your Data" program.[6] The data include 5-minute interval observations of thermostat settings and the amount of time the furnace fan was running. In addition, a small amount of information about the household is available, including the location up to city and state, the number of occupants, the size, age, and number of floors

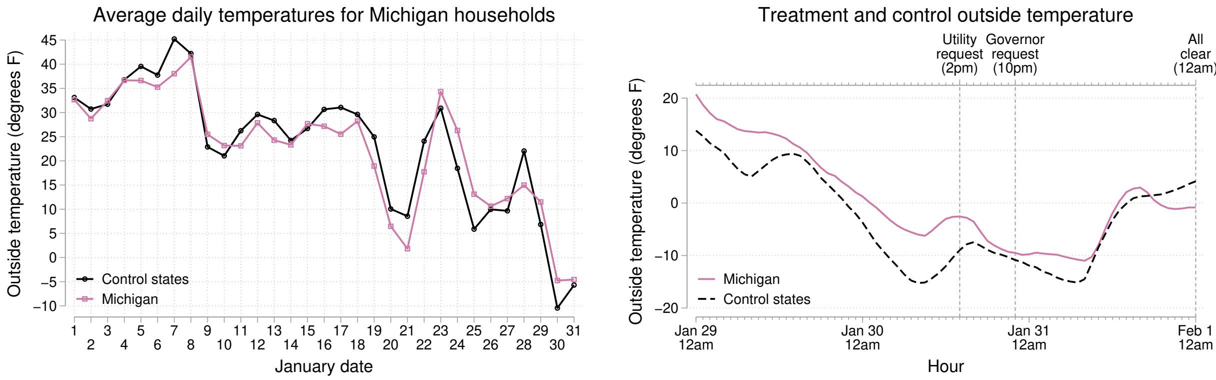
[6]This paper is among a few others studying the effects of smart thermostats or using smart thermostat data. Ge and Ho (2019) study how households change the thermostat in response to warm and cold weather and assess the degree of habit formation in thermostat settings. A working paper by Brandon et al. (2021) find that smart thermostats alone do not result in energy savings, partially due to users overriding smart thermostat algorithms. Another working paper by Blonz et al. (2021) studies an energy-efficiency program implemented by Ecobee that automatically reduces thermostat settings during peak pricing periods.

of the home, and when the smart thermostat was first connected. The outdoor temperature and relative humidity for each household is included, but we find this data is often missing, so we replace it with hourly outdoor temperature and humidity at the city level purchased from Visual Crossing.[7] Consumers Energy only serves households in Michigan. We limit the sample of households to those in Michigan and the surrounding four states for controls: Ohio, Indiana, Illinois, and Wisconsin.[8] 99.89 percent of sample households heat with natural gas, compared to 75 percent of population households in Michigan. We include all observations between January 1st and February 7th, 2019. There are 2,700 households from Michigan and 7,953 control households in the final sample.

It is possible that the households in our sample responded to the emergency request differently than the general population due to selection into smart thermostat ownership and the Donate-Your-Data program. On observable characteristics, the Ecobee Donate-Your-Data households are comparable to the average household in the nationally representative Residential Energy Consumption Survey sample, though the Ecobee households have slightly more members (Meier et al., 2019). Our primary selection concern is that Ecobee households who join the Donate-Your-Data program may be more likely to contribute to other public goods and therefore more likely to comply with the emergency request. Another concern we have is that the Ecobee smart thermostat may make compliance with the request easier than compliance using a conventional thermostat because Ecobee thermostats can be controlled remotely via an app. While these issues are not a problem for our research design because treatment and control households are the same (i.e., our research design is internally valid), it may be that our estimated treatment effects overstate the response of the average household. However, when we compare our average treatment effects to the reduction in consumption estimated by the utility using aggregate data, we find that our estimates are similar but slightly smaller in magnitude, which is the opposite of what we would expect if selection was

[7] Regression coefficients do not substantially change based upon which weather variables are used.

[8] Households in the Upper Peninsula of Michigan were excluded from the analysis because these households are on a separate natural gas network and it is unclear whether they were treated or were controls. The Upper Peninsula accounts for about 3% of the population of Michigan; dropped households represent 1% of the Michigan sample in the data.



(a) (b)

Figure 2: (a): Sample mean daily temperatures for Michigan and control households in January 2019. January 20-21 are used as a "placebo" event for the January 30-31 polar vortex and emergency request. (b): Sample mean hourly outdoor temperatures for Michigan and control households in the hours before and during the emergency.

substantially affecting the estimates. This alleviates our concerns about selection, but we conservatively interpret our average treatment effect estimates as an upper bound.

For computational tractability, we aggregate the data into hourly time intervals, resulting in 2.4 million household-hourly observations in Michigan and 7.1 million household-hourly observations in the controls.[9] Table 1 displays summary statistics for the treatment and control groups. Due to the large sample size, most differences in means are statistically significant, but are not practically meaningful. The primary differences we see are that sample homes in Michigan are slightly smaller and have fewer occupants on average. During January 1 - 29, Michigan and the control states experienced average temperatures around 24 and 25°F. The outside average temperature in both treatments and controls dropped to just under 0°F during the event. From January 1st through 29th, Michigan household thermostat settings were 0.5°F lower than the control household thermostat settings. In the hours before the first appeal to lower thermostats, the gap in thermostat settings had shrunk to 0.37°F. After households were asked to reduce the thermostat, the gap increased to 1.32°F.

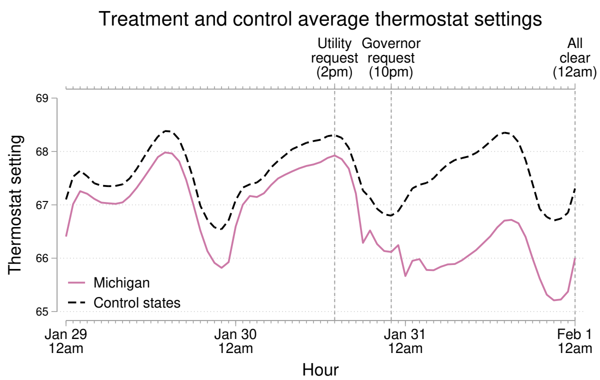
[9]We compute the average thermostat setting and minutes the furnace was running. Regressions using data at the half-hour level do not substantially change the coefficient estimates.

Our research design compares outcomes in Michigan to those in control states where there were no appeals to reduce natural gas consumption. We consider three outcomes: the thermostat setting, a binary variable equal to one if the thermostat setting is below 65°F, and the amount of time the furnace fan ran during the hour. Standard furnaces run at essentially one speed.[10] When the thermostat setting is reduced, the home cools to the new setting and the furnace does not run, saving energy. When the indoor temperature goes below the new thermostat setting, the furnace runs again at full speed for a short period to maintain the indoor temperature. Thus, furnace fan running time is our best proxy for natural gas consumption (Meier et al., 2019). Consumers Energy shared daily aggregate natural gas consumption and forecasts of expected consumption from their internal forecasting model, which we use to construct an estimate of total demand response from all sources.

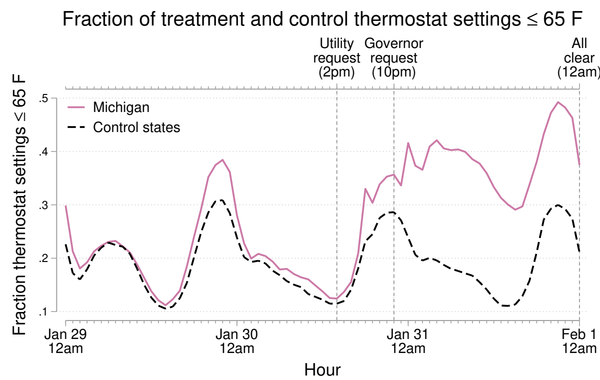
Households in the control states (the "Great Lakes states": Ohio, Indiana, Illinois, and Wisconsin) have weather patterns and housing stocks similar to Michigan. Furthermore, these states also experienced extreme cold during the polar vortex event. Figure 2a plots daily average temperatures during January for Michigan and control states, showing the polar vortex event at the end of the month in addition to a similar cold wave on January 20th and 21st that we study in a placebo exercise in the appendix. Figure 2b plots the average hourly outdoor temperature before and during the emergency, demonstrating that both treatment and control groups experienced similar conditions during the polar vortex.

Because we observe treatment and control households before and during the event, the difference-in-differences framework is a natural candidate to estimate the effect of the emergency request on thermostat settings. The key assumption needed in a difference-in-differences design is a parallel trends assumption in the evolution of the potential untreated outcome. To demonstrate the validity of the difference-in-differences assumption, we show that our outcome variables exhibit parallel trends prior to the event. Figure 3 plots sample average hourly thermostat settings, the fraction of households with thermostat settings below 65°F, and the number of minutes the fan was running in Michigan and the control

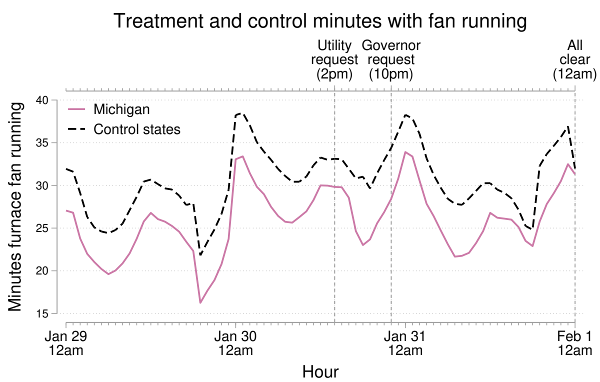
[10] Two-stage furnaces can run at full-speed and half-speed depending on the scenario to reduce energy use and ramping costs.



(a)



(b)



(c)

Figure 3: Sample average values of the outcome variables for treatment and control households, January 29 - February 1. Panel (a) plots average thermostat settings, panel (b) plots the fraction of households with thermostat settings below 65°F, and panel (c) plots the average furnace fan run time in each hour.

states from January 29th through February 1st. The first vertical dashed line indicates when the utility company first broadcast a request to residential customers to reduce natural gas consumption by reducing thermostats, the second vertical dashed line indicates when the Governor broadcast the emergency appeal, and the final vertical dashed line indicates the all-clear time.

Prior to the event, the treatment and control thermostat settings, fraction of households with thermostat settings below 65°F, and the number of minutes the fan was running in Michigan exhibit roughly parallel trends without conditioning on covariates. When the event begins, a take-up lag can be observed where households have either not received the message or are not home and able to respond. A few hours after the event begins, the average thermostat setting in Michigan breaks trend and significantly decreases. Despite the Governor's request that households reduce thermostats to 65°F or lower, the average observed smart thermostat setting in Michigan is above 65°F during the entire event. It appears that compliance with the request does not begin until after the Governor's appeal. The differences in furnace fan running time after the request are more difficult to detect visually than thermostat setting and fraction of compliant households.

We supplement the smart thermostat data with data on gubernatorial election outcomes for each state's most recent election by county obtained from the Voting and Elections Collection maintained by the CQ Press (2019). Because the emergency requests were eventually taken up by the Governor of Michigan, we hypothesize that compliance varies based upon support of the Governor. To test this hypothesis, we use a triple-differences strategy, estimating treatment effects at the county level, using support for the Democratic party in the most recent gubernatorial election as the third difference. We include household-level controls available in the Ecobee data as well as demographic controls at the county level obtained from the American Community Survey (US Census Bureau, 2019).

4 Empirical analysis and results

We compare differences in outcomes between households in Michigan and surrounding states before and after the emergency requests. We consider three outcomes. The household's thermostat setting is a continuous measure of the household's compliance with the emergency request and encompasses the thermal discomfort that the household incurred to contribute to the public good. The second outcome is a binary variable equal to one when the thermostat setting is below 65 degrees Fahrenheit. This binary variable captures whether households complied with the request to the letter. The final outcome variable, number of minutes the fan ran during the hour, is the best proxy available for the amount of energy conserved.

The analysis is divided into four subsections. We begin with a simple pre/post difference-in-differences framework to estimate the average treatment effect of the program. We then move to an event-study framework that allows for dynamic effects by hour as word of the emergency reached more households. Next, we test whether support for the Governor of Michigan affected compliance rates. Finally, we study how the phrasing of the emergency request around 65°F influenced household behavior.

4.1 Average treatment effect estimates

The first set of regressions we consider are two-way fixed effects difference in differences specifications on all January 2019 observations. We consider outcomes $Y_{i,t}$ and code a binary variable $D_{i,t} = 1$ for all Michigan observations beginning January 30th at 2:00 pm and zero beforehand. Our preferred specification takes the following form:

$$Y_{i,t} = \alpha_i + \lambda_t + \beta D_{i,t} + \gamma X_{i,t} + \varepsilon_{i,t}, \tag{1}$$

where α_i are household fixed effects, λ_t are hour-of-sample indicator variables, $X_{i,t}$ are controls for outside temperature and humidity, and $\varepsilon_{i,t}$ is mean-zero heterogeneity. Equation 1 is a two-way fixed-effects specification. The ordinary-least-squares estimate $\hat{\beta}$ is a difference-in-differences estimate that identifies the causal average treatment effect on the treated under

Table 2: Estimates of the regressions from equation 1.

Two-way fixed effects regressions

	(1)	(2)	(3)
VARIABLES	Thermostat setting	Compliance	Fan run time
Michigan x Post	-0.855**	0.114**	-1.952**
	(0.047)	(0.006)	(0.315)
Constant	67.104**	0.223**	25.225**
	(0.030)	(0.003)	(0.147)
Observations	7,604,174	7,604,174	7,604,190
R-squared	0.698	0.465	0.628
Weather	YES	YES	YES
Household FE	YES	YES	YES
Time FE	YES	YES	YES

Robust standard errors clustered at the city level.
** $p<0.01$, * $p<0.05$

the standard parallel trends, no spillovers, and strict exogeneity assumptions.

Table 2 presents the coefficient estimates of the difference-in-differences regressions for each of the three outcome variables: thermostat setting, a binary variable for setting the thermostat below 65 degrees Fahrenheit, and the number of minutes the furnace fan ran. We cluster the standard errors at the city level.[11] When the thermostat setting is the outcome variable (column 1), the coefficient on $D_{i,t}$ is an estimate of the average treatment effect on the treated and is the mean difference in thermostat settings for Michigan and control states before and after the treatment. We estimate a reduction of 0.86°F after the emergency request for Michigan households relative to neighbor state households. This reduction is about 0.22 standard deviations in the thermostat setting from January 1 - 29.

Column 2 presents estimates using an indicator variable for having the thermostat below 65°F as the outcome variable, which we interpret as full compliance with the request. Given that at any time, some fraction of Michigan households would already have thermo-

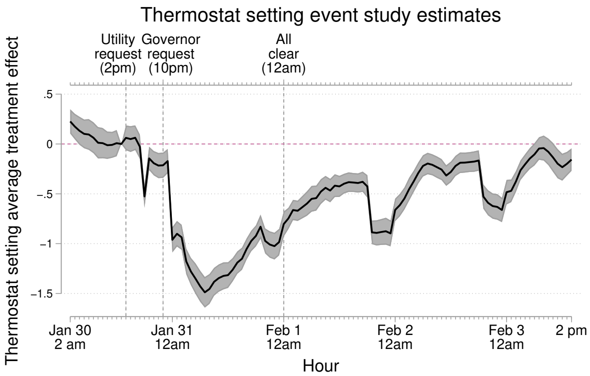
[11] Ideally we would cluster at the state level given it is the level of treatment (Abadie et al., 2017), but given that there are only five state clusters, it is unlikely that the cluster-robust standard error estimators will converge (Cameron et al., 2008). Instead, we cluster at a lower level of aggregation to increase the number of clusters to ensure convergence. The main estimates are still statistically significant even when clustered at the state level, but we do not report these standard errors.

stat settings at 65°F or below, the difference-in-differences estimate accounts for this by differencing out the within-household and within-time average incidental compliance. The coefficient on $D_{i,t}$ is an estimate of the additional fraction of households induced to set the thermostat below 65°F. We estimate an 11.4 percentage point increase in the fraction of households with thermostat settings below 65°F for Michigan households relative to neighbor state households. Using the "constant" term reported in the two-way fixed effects estimates, the expected number of households in Michigan that already would have had thermostat settings below 65 degrees was 22 percent, which we refer to as the incidental compliance.

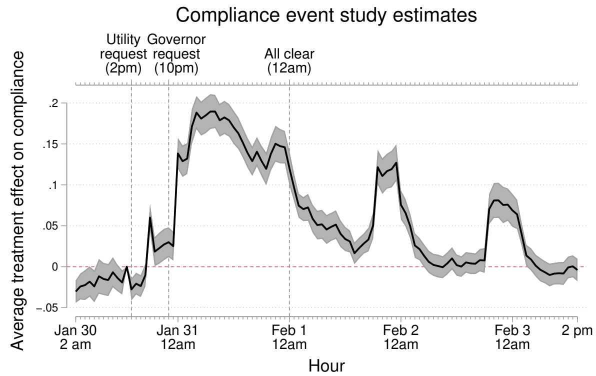
Finally, column 3 presents the coefficient estimates of the difference-in-differences regressions using furnace fan run time as the dependent variable, which is the closest proxy to energy consumption in the smart thermostat data. We estimate a 2 minute per hour average reduction in furnace fan run time for Michigan households relative to neighbor state households. Relative to the conditional mean furnace fan run time for Michigan households during the emergency, this is an 8.2 percent decrease. Given the lack of natural gas consumption data at the household level, this is the best estimate of the amount of natural gas savings caused by the emergency request.[12] Using aggregate daily consumption data provided by Consumers Energy, the total reduction in natural gas consumption from all sources was about 10 percent, which is in line with our estimates.

We conduct a series of robustness checks and a placebo test in appendix sections A and B. The robustness checks test the sensitivity of the average treatment effects to alternative difference-in-differences specifications, omitting households who join the sample late or leave early (i.e., using a balanced panel), allowing for spillovers to counties bordering Michigan, and to omitting households enrolled in Ecobee's "Eco+" energy efficiency program.[13] We find that the estimated effects do not change substantially. The placebo test analyzes a cold wave in Michigan that occurred ten days earlier on January 20-21, 2019, where temperatures dropped by a similar magnitude. We find that Michigan's heating behavior remains parallel to the control households during this placebo event and regression 1 yields estimates of zero

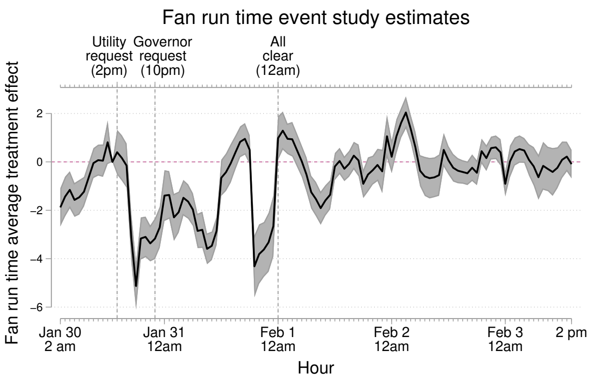
[12]Natural gas consumption at the household level is measured at the monthly level by the utility.
[13]this program is studied in a working paper by Blonz et al. (2021).



(a)



(b)



(c)

Figure 4: Event-study coefficients estimated using regression equation 2 with (a) thermostat setting, (b) compliance, and (c) minutes of furnace fan run time as the dependent variables. 95 percent confidence intervals constructed from standard errors cluster-robust to heteroskedasticity.

using the placebo treatment, suggesting that our findings are not an artifact of differential responses by Michigan households to cold waves. In our regular specifications, this cold wave is included in the data and thus serves as a control, lending credibility to the research design.

4.2 Event-study estimates

Given the repeated requests for reductions in thermostat settings over time, we next account for a dynamic response in an event-study framework. We estimate a two-way fixed effects regression using the following specification:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{k=-12}^{-2} \beta_k^{lead} \mathbf{1}[k = t - g] + \sum_{k=0}^{96} \beta_k^{lag} \mathbf{1}[k = t - g] + \gamma X_{i,t} + \varepsilon_{i,t}, \qquad (2)$$

where g is the time period the utility made its first emergency request to households. Thus, we estimate 11 lead coefficients and 97 lag coefficients to include a half day of pre-trends and four days of dynamic treatment effects.[14] We hypothesize that prior to the end of working hours, household responses will be muted and that the largest responses will occur after the Governor's use of the emergency text message alert at 10:30 pm. Further, we suspect that the treatment effect persisted after the "all clear" time given barriers to receiving the all-clear message or adjusting the thermostat (e.g., if the home is vacant or all occupants are sleeping).

Figure 4 plots the dynamic treatment effects estimated using the event-study regressions specified in equation 2 using thermostat setting, compliance with the request, and fan run time as outcome variables. The first finding of note is that the Governor's alert was essential to increasing compliance. Averaging over the eight hours of event-study coefficients prior to the Governor's alert, the utility's emergency request only resulted in an average additional compliance rate of 0.5 percent (with a peak rate of about 6 percent), resulting in an average thermostat reduction of just 0.12°F. Following the Governor's alert, the average additional compliance rate was 14.7 percent (with a peak of about 18 percent), resulting in an average thermostat reduction of 1.08°F.

While the initial lukewarm response to the utility's emergency appeal may have been caused by households not being at home to change the thermostat setting, the average additional compliance only peaked at 5% before the Governor's request. Given the utility's actions of sending emails to customers, posting on social media, and reaching out to traditional news media, we do not think the lack of responsiveness was solely due to a lack of reach. Instead, it is likely that households did not take the request seriously until it became clear that there was a true emergency. The additional authority of the Governor and the repeated request to reduce thermostat settings likely increased the salience of the request, inducing additional compliance.

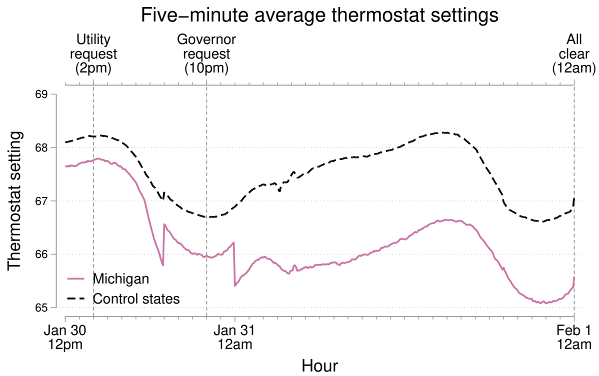
[14] Choosing greater or fewer leads and lags does not substantially change the coefficient estimates, but increases computational cost. We chose the window to allow us to observe when thermostat settings returned to a normal level after the event.

After the emergency request and before the "all clear," thermostat settings begin to trend upward and compliance begins to fall. Sustained compliance is likely increasingly costly, so participation rates decline over time. This trend may also be due to households choosing low thermostat settings when sleeping and leaving the home for work in the morning. Upon returning from work, households may increase the thermostat setting to a slightly higher level. This behavior is similar to the "backsliding" dynamic reported by Allcott and Rogers (2014) in which households conserve electricity after receiving a home energy report, but the effect lessens over time. To sustain high levels of compliance in an emergency, repeated requests are likely necessary.

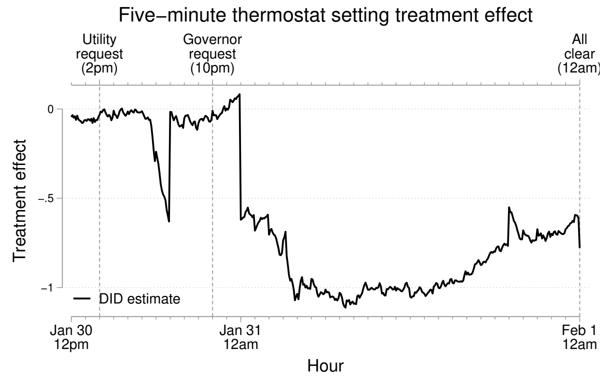
Another interesting finding is that the effect persists after the "all clear." Not only does it take time for the average thermostat setting to return to normal, but there is also a dip in thermostat settings around 8 pm. This dip suggests that households had programmed their thermostats to reduce the temperature setting for the nighttime in keeping with the emergency request and had not yet re-programmed them after the all clear. This result is consistent with previous work that finds that changes in thermostat settings in response to a cold or hot period tend to persist after the cold or hot period ends (Ge and Ho, 2019).

In addition, one can see that the reductions in furnace fan running time are only transitory. Because furnaces essentially run at one speed, reducing the thermostat at night or when out of the home will reduce energy use while the home cools, but upon increasing the thermostat, the furnace will need to run again and incur a ramp-up cost to increase the temperature. We can see in the fan run time event study estimates that on January 31st, there were savings during the early morning and day, but when households returned home in the evening that furnaces had to run at essentially full intensity to warm the home again. After the all clear, the estimates return to mean zero more quickly than the thermostat setting and compliance rate estimates.

We supplement the hourly event study with a graphical analysis of the five-minute thermostat-setting data. In figure 5a, we plot five-minute thermostat setting data for Michigan and the control states between 12:00 pm on January 30th and 11:59 pm on January 31st.



(a)



(b)

Figure 5: (a): Five-minute sample mean thermostat settings for Michigan and control households from January 30th 12 pm - January 31st 11:59 pm. (b): Five-minute difference-in-differences estimate.

In figure 5b, we plot a difference-in-differences estimate of the treatment effect, which we construct as the difference between five-minute thermostat setting for Michigan and control states during the event minus the average difference in same hour-of-day and day-of-week five-minute thermostat settings before the event. In these figures, one can clearly see Michigan households begin responding to the requests beginning around 5:00 pm on January 30th, likely as households returned home from work. At 7:00 pm, there is a discrete increase in thermostat settings in both Michigan and the control states. This change is exactly at the top of the hour and corresponds with a common time that households program into the thermostat to automatically adjust the indoor temperature. The previously programmed thermostat setting appears to have overridden household compliance efforts in many households. In the five-minute period beginning 12:00 am on January 31st, we observe a discrete decline in Michigan thermostat settings. Thus, it appears Michigan households used the programmable thermostat features to comply with the request when going to bed. Notably, we do not see a discrete change directly after the Governor's request.

In appendix section B.2, we replicate the five-minute analysis during the placebo cold wave. The difference-in-differences estimates are zero throughout most of the placebo period (other than in a few spurious five-minute periods), lending credibility to the difference-in-differences estimates in figure 5b. In addition, we see similar discrete increases in temperature for Michigan and the control states in the five-minute placebo periods beginning at 7:00 pm and 12:00 am. One important difference is that during the polar vortex event, the discrete change in Michigan at 12:00 am is a decrease, while in the placebo event it is an increase. The placebo event confirms that these changes at the top of the hour likely reflect user-programmed thermostat changes.

The results from the five-minute analysis suggest that automation comes with potential risks and benefits. The increase in Michigan thermostat settings beginning at 7:00 pm show how user-programmed changes can override household attempts at compliance. On the other hand, the decrease in Michigan thermostat settings beginning at 12:00 am show how user-programmed changes can be used to comply with emergency directives. Future requests for thermostat reductions may be more effective if they include short instructions regarding overriding programmable thermostats. In addition, as AI-enhanced thermostat-control software increases the prevalence and varieties of automation, developers and utilities may consider automated emergency overrides and programs.

4.3 Heterogeneity analysis

Next, we analyze support for the Michigan Governor and the effect of the reference point on household behavior in a triple-differences framework. Our approach analyzes both forms of heterogeneity in the same estimating equation with additional controls to account for the possibility that county-level support for the Governor is correlated with baseline thermostat setting or other demographic factors. Thus, we discuss how we define baseline thermostat setting and support for the Governor separately before presenting the joint estimating equation.

Households whose thermostats would have been at 65°F or lower essentially received information that they were already keeping the thermostat low enough and may have felt that they did not need to reduce the thermostat further.[15] Furthermore, we hypothesize

[15]This analysis generally treats larger reductions as welfare-improving, but we note that thermostat set-

that the distance from the reference point may also effect household behavior. The higher a household's normal thermostat setting, the larger the requested reduction in temperature. This means that for compliers with a high baseline thermostat setting, the treatment effect will be larger by construction; however, the deviation from normal behavior is greater and so we might expect this to discourage compliance. To test this hypothesis, we estimate the effect of the emergency request allowing for different responses by expected baseline thermostat settings. We construct a non-parametric estimate of baseline expected thermostat setting $\hat{T}_{i,t}$ for each household by calculating the household's sample average thermostat setting for each day-of-week and hour-of-day combination from the pre-treatment period. Denote $\mathcal{T} = \{[0, 59), [59, 61), [61, 63), ..., [73, 75), [75, 100]\}$ as the collection of 2-degree intervals from 59°F to 75°F with binned endpoints for higher and lower temperatures, and $b \in \mathcal{T}$ the interval with upper bound b. We interact indicator variables for belonging in each interval with the treatment variable to create a third difference and estimate heterogeneous effects by baseline temperature category.

In the same regression, we analyze heterogeneity by political support for the Governor. The data on gubernatorial election returns is at the county level. In the 2018 election, the distribution of the Michigan Governor's county vote share ranged from 31 percent to 73 percent. We expect the effect of political support to be non-linear so we create 5-percentile indicator variables between 30 and 75 percent. Denote $\mathcal{P} = \{[30, 40), [40, 45), [45, 50), ..., [70, 75]\}$ as the collection of a 10-percentile interval between 30 and 40 and 5-percentile intervals between 40 and 75, and $a \in \mathcal{P}$ the interval with upper bound a.[16] Similarly to the baseline thermostat setting, we interact indicator variables for belonging in each interval with the treatment variable.

tings that are too low increase the risk of frozen pipes.

[16]We combined the 30-35 and 35-40 percentile intervals because only 0.38 percent of households lived in counties with a Democratic Party vote share between 30 and 35 percent, which lead to extremely imprecise estimates.

Thus, our estimating equation is

$$
\begin{aligned}
Y_{i,t} = \; & \alpha_i + \lambda_t + \sum_{b \in \mathcal{T}} \beta_b D_{i,t} \times \mathbf{1}[\hat{T}_{i,t} \in b] + \delta_b \mathbf{1}[\hat{T}_{i,t} \in b] + \sum_{a \in \mathcal{P}} \beta_a D_{i,t} \times \mathbf{1}[P_{county} \in a] \\
& + \gamma_1 D_{i,t} \times Z_i + \gamma_2 X_{i,t} + \varepsilon_{i,t},
\end{aligned} \tag{3}
$$

where Z_i is a vector of controls for household-level characteristics available in the smart thermostat data as well as county-level demographics to address correlation between county vote share and demographics.[17] Given our hypotheses, we expect compliance to fall with an increased baseline expected thermostat setting. Thus, we expect β_b to be lower for higher levels of b. It is possible for the average treatment effect to either increase or decrease with a higher baseline expected thermostat setting. Moreover, it is possible that very cold baseline households may increase thermostat setting when introduced to the reference level of 65, thus we expect β_b to be zero or positive for $b \leq 65$.[18] For the thermostat setting and compliance outcome variables, we hypothesize that the coefficients on the interaction with vote share β_a will be increasing as Democratic vote share increases and that the opposite will be true for the fan running outcome variable regression, indicating that the appeal was more effective for households in counties that supported the Governor's election.

Table 3 displays the estimates of equation 3 for each outcome variable. The standard errors are cluster-bootstrapped to incorporate the uncertainty due to sampling error from estimating the baseline thermostat setting.[19] We discuss the vote share and baseline thermostat setting results separately in the following sections.

[17] From the smart thermostat data, we include square feet of the home, number of occupants, whether the home is detached or an apartment, and the age of the home. From the ACS, we include county level median income, median age, population, fraction male, fraction white, fraction with a high school education or more, and fraction of non-US citizen residents.

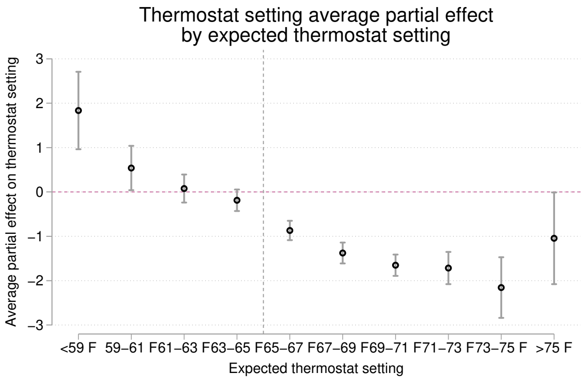
[18] This effect is seen in the social-comparison literature where if a household receives information that they are consuming less than the average they may decide they were being too conservative and increase consumption.

[19] The bootstrap procedure first draws observations with replacement from within household, day of week, and hour of day strata to estimate 100 different baseline temperatures for each household, day of week, and hour of day combination. It then samples from this empirical distribution and draws 100 bootstrap samples clustered at the city level. Ultimately, these standard errors differ very little from clustered standard errors that ignore the uncertainty from the first-stage estimation.

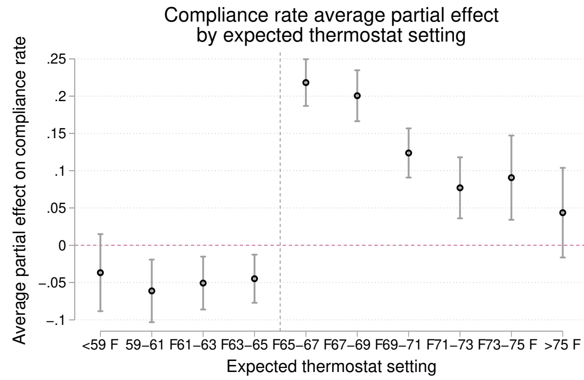
Table 3: Results from the estimation of equation 3. Standard errors cluster-bootstrapped to incorporate the sampling error from estimation of the baseline thermostat setting.

	(1) Thermostat setting	(2) Compliance	(3) Fan run time
40-45% Democrat X Treatment	-0.10	0.045*	0.050
	(0.19)	(0.019)	(0.81)
45-50% Democrat X Treatment	-0.25	0.064*	1.37
	(0.26)	(0.028)	(1.05)
50-55% Democrat X Treatment	-0.60**	0.073**	0.97
	(0.21)	(0.024)	(0.90)
55-60% Democrat X Treatment	-0.42	0.078*	3.36*
	(0.29)	(0.033)	(1.55)
60-65% Democrat X Treatment	-0.85**	0.14**	1.93
	(0.32)	(0.043)	(1.62)
65-70% Democrat X Treatment	-0.62	0.070	3.10
	(0.44)	(0.058)	(1.69)
70-75% Democrat X Treatment	-0.95*	0.13**	3.65*
	(0.38)	(0.045)	(1.73)
59 F or lower expected X Treatment	2.02**	0.0081	2.26**
	(0.43)	(0.025)	(0.71)
59-61 F expected X Treatment	0.73**	-0.016	0.88
	(0.22)	(0.018)	(0.64)
61-63 F expected X Treatment	0.26**	-0.0058	0.51
	(0.092)	(0.013)	(0.43)
65-67 F expected X Treatment	-0.68**	0.26**	-1.07**
	(0.049)	(0.012)	(0.26)
67-69 F expected X Treatment	-1.19**	0.25**	-1.26**
	(0.063)	(0.013)	(0.29)
69-71 F expected X Treatment	-1.46**	0.17**	-1.06**
	(0.080)	(0.014)	(0.34)
71-73 F expected X Treatment	-1.53**	0.12**	-1.71**
	(0.17)	(0.018)	(0.45)
73-75 F expected X Treatment	-1.97**	0.14**	-0.93
	(0.30)	(0.024)	(0.99)
Higher than 75 F expected X Treatment	-0.86	0.089**	-3.57*
	(0.50)	(0.027)	(1.69)
Observations	7,051,137	7,051,137	7,051,159
R-squared	0.78	0.61	0.63
FE	YES	YES	YES
Hour	YES	YES	YES
Controls	YES	YES	YES
Expected thermostat level	YES	YES	YES

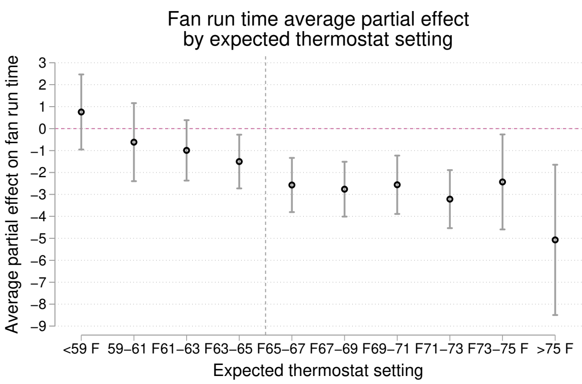
Standard errors cluster-bootstrapped at the city level.
** p<0.01, * p<0.05



(a)



(b)



(c)

Figure 6: Effect of the emergency request by expected thermostat setting estimated using equation 3 with (a) thermostat setting, (b) compliance, and (c) minutes of furnace fan run time as the dependent variables. 95 percent confidence intervals constructed from bootstrapped standard errors that account for first-stage estimation of expected thermostat setting and are cluster-robust to heteroskedasticity.

4.3.1 Reference point effect

The coefficients on the interaction between expected thermostat setting and treatment in table 3 are the difference in the average treatment effect by expected thermostat setting relative to households in hours with an expected thermostat setting of 63-65°F (the omitted category). We find that households in hours with baseline thermostat settings below 65°F were less responsive to the appeal. As the baseline thermostat setting increases above 65°F, the magnitude of the response increases and then decreases at higher temperatures. Compliance with the emergency request falls as the baseline thermostat setting increases above the requested level. The estimates for the fan run time variable show that for the coldest baseline thermostat setting, fan running times increased relative to the base category. For baseline thermostat settings above 65°F, fan run time decreased relative to the base category, with the largest effect (though not statistically significant) for the highest baseline thermostat settings.

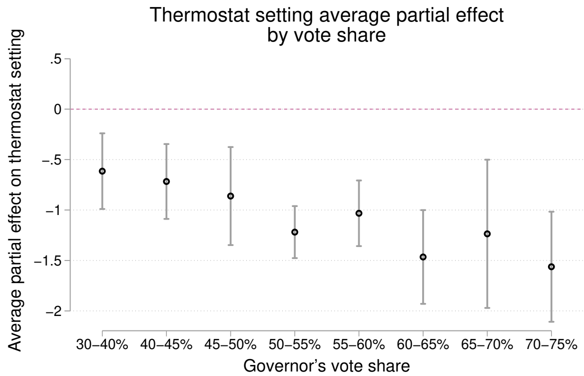
One concern we had was whether these estimates were an artifact of statistical mean reversion rather than a meaningful pattern.[20] To test this alternative hypothesis, we estimate equation 3 using the placebo cold wave event. Appendix section B.3 displays the results of the placebo analysis. The estimates from the placebo analysis are the same sign as the estimates from the polar vortex, but the magnitude of the estimated coefficients in the placebo analysis are almost all two to five times smaller than during the polar vortex. Furthermore, the placebo estimates are relatively flat as the baseline category gets further away from 65°F. Thus, we conclude that mean reversion may play a small role in the main heterogeneity estimates but the effect is not large enough to alter our conclusions in this section.

Figure 6 displays the average partial effects of treatment on the outcome variables by baseline thermostat setting.[21] When the expected thermostat setting is less than 65°F, the appeal does not decrease the thermostat setting and is likely to increase the thermostat setting in the coldest homes. On average, when the expected thermostat setting is below 65°F, the appeal corresponds with households increasing the thermostat above 65°F. These perverse effects are consistent with the appeal anchoring low thermostat settings to the norm of 65°F. Alternatively, the appeal may have increased the sense of danger, causing households to stay home when they typically would have left for work and reduced the thermostat setting.

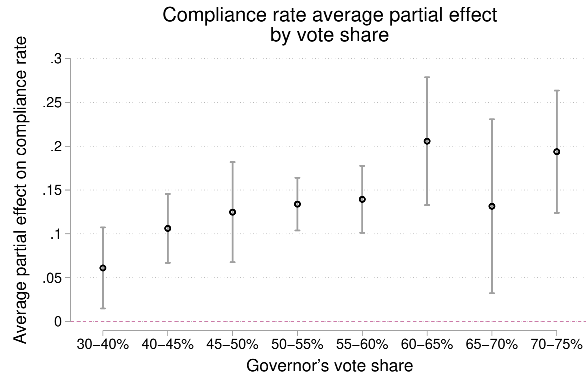
Households were also less likely to fully comply when the reference level of 65°F appeared out of reach. This suggests that a reference level can induce larger contributions of effort for households near the reference level, but it discourages effort for households far away from that reference level. For households with expected thermostat settings above 75°F, the average partial effect on thermostat setting and compliance rate are the lowest of those above the reference level; however, the average partial effect on fan run time was the largest.

[20] That is, do our estimates merely reflect that households with high or low temperatures in the past are mechanically more likely to have average temperatures when measured later?

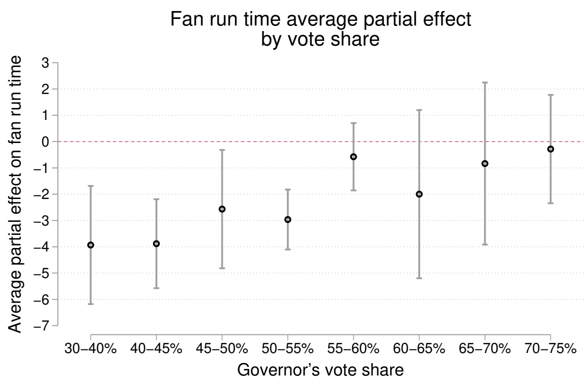
[21] Formally, the average partial effect of treatment by baseline thermostat setting $\hat{T}_{i,t} \in b$ is $E[Y_{i,t}|W_{i,t}, \hat{T}_{i,t} \in b, D_{i,t} = 1] - E[Y_{i,t}|W_{i,t}, \hat{T}_{i,t} \in b, D_{i,t} = 0]$, where $W_{i,t}$ is a vector of all other control variables in equation 3.



(a)



(b)



(c)

Figure 7: Effect of the emergency request by Governor's vote share estimated using equation 3 with (a) thermostat setting, (b) compliance, and (c) minutes of furnace fan run time as the dependent variables. 95 percent confidence intervals constructed from standard errors cluster-robust to heteroskedasticity.

This suggests that those households that did comply generated substantial energy savings, although the heterogeneity of this effect leads to a wide confidence interval.

4.3.2 Vote share effect

The coefficients on the interaction between county vote share and treatment in table 3 are the difference in the average treatment effect for households in counties with a given level of support for the Governor relative to households in counties where the Governor's vote share was 30-40% (the omitted category).[22] The estimates show that the average thermostat reduction and compliance rate is higher in all counties relative to those with the lowest Governor's vote share. In addition, the treatment effects are generally increasing in magnitude as the Governor's vote share increases. The results indicate that households in

[22] Appendix section 4.3 replicates this analysis using the placebo cold wave and does not find the same patterns in the estimated coefficients.

the counties that supported the Governor the most reduced thermostats by up to one degree Fahrenheit more on average and had a compliance rate 13 percentage points higher relative to the least supportive county. Despite this, the estimates of the effect on fan run time are the opposite of the expected sign and are imprecise with most confidence intervals containing zero. Given that fan run time is a function of underlying energy efficiency of the furnace and home, we believe that the fan estimates reflect unobserved differences in energy efficiency that are correlated with political affiliation.

Figure 7 displays average partial effects of treatment on the outcome variables by Governor's vote share.[23] The average partial effect on thermostat setting and compliance is increasing in the Governor's vote share. On average, the appeal induced about a 6 percent compliance rate in the counties most opposed to the Governor and a 19 percent compliance rate in the counties most in support of the Governor. While this effect is large, it does not outweigh the increased compliance rates generated by the Governor's amplification of the appeal on social media and via the emergency text alert system.

Thus, we conclude that support for the Governor is correlated with a stronger response to the public appeal, although the implications for energy use are unclear. We caution against interpreting these estimates causally, but our findings are consistent with distrust arising out of affective political polarization. In the increasingly polarized political environment of the United States (Iyengar et al., 2019), a public appeal may be met with defiance rather than compliance. An alternative explanation is that political ideology may be correlated with willingness to contribute to public goods or thermostat setting behavior more broadly. Given our inability to distinguish the effect of polarization from political ideology, we cannot reject either explanation.

[23]The average partial effect of treatment by baseline county-level vote share $\hat{T}_{i,t} \in b$ is $E[Y_{i,t}|W_{i,t}, \hat{T}_{i,t} \in b, D_{i,t} = 1] - E[Y_{i,t}|W_{i,t}, P_{county} \in a, D_{i,t} = 0]$, where $W_{i,t}$ is a vector of all other control variables in equation 3.

5 Conclusion

This paper studies an acute natural gas shortage during the 2019 polar vortex in Michigan. During near-record low temperatures, a fire at a compressor plant resulted in natural gas demand that nearly outpaced supply. In response, the utility requested households to voluntarily conserve natural gas, and the Governor of Michigan subsequently issued an emergency text alert that requested households voluntarily reduce thermostat settings to 65°F.

We use smart thermostat data to analyze consumer responses to the emergency request, finding robust evidence of voluntary compliance with the request. Using a difference-in-differences strategy with four control states, we obtain estimates of the average treatment effect on the treated. On average, households lowered their thermostats by 0.86°F, roughly a 20% reduction of the typical variation in the average thermostat setting. 11 percent of households complied with the request fully by reducing their thermostats to below 65 degrees Fahrenheit, while 24 percent of household thermostat settings were already below the threshold. Finally, we find evidence that the emergency request reduced furnace fan run times (our best proxy for natural gas consumption) by 2 minutes per hour; an 8.2 percent decrease. This reduction is slightly smaller than the 10 percent reduction in all consumption of natural gas we calculate using aggregate consumption data provided by the utility.[24] Our estimate is comparable with reductions in energy consumption observed in field experiments that use moral suasion to induce conservation, but falls short of field experiments that use price incentives to induce conservation (see e.g., (Ito et al., 2018)).

Our analysis highlights the importance of wide-reaching emergency messaging for governments and utilities. An event study analysis reveals that prior to the Governor's announcement, the utility's emergency request only induced 5 percent of households to reduce thermostat settings to 65°F or less. After the Governor's amplification of the emergency request, the fraction of households in compliance with the request nearly quadrupled to 18 percent at its height. The emergency directives communicated via social media and news

[24] Aggregate consumption data includes residential, commercial and industrial consumption.

media suffered from low visibility and were likely lost in the large amount of other content on these platforms. As the emergency progressed, compliance with the request waned as households likely found it increasingly costly to maintain low thermostat settings. In addition, we find evidence of persistence in low thermostat settings in the day after the emergency. These habits appeared to be driven by programmed thermostat settings left in place by households.

The particular phrasing of the emergency request around the reference point of 65°F played a large role in determining household behavior. We identify three perverse effects of the reference point. First, households that typically heat at 65°F or lower did not reduce thermostats in response to the emergency request. Second, those that typically have the lowest thermostat settings even increased the thermostat setting after receiving the request. Third, those with the highest thermostat settings were less likely to comply with the request. For households with thermostat settings typically above 65°F, the average treatment effect at first increases and then decreases with distance from the reference point. Mechanically, households with a higher thermostat setting must reduce it more to comply with the request; however, the cost of doing so increases, discouraging compliance for households that typically have the highest thermostat settings.

Setting a more aggressive target trades off a larger effect of compliance with the cost of compliance, which increases defiance. This suggests that a particular reference point may induce a larger emergency response. While this natural experiment did not provide the necessary variation to determine the highest-impact reference points in this context, these could be determined via purposeful experimentation. Furthermore, smart thermostat technology offers the possibility for targeted requests and automated compliance. Alternatively, requests for a uniform reduction in thermostat setting may induce broader compliance than a uniform compliance goal because they avoid the problems driven by users in the tails of the energy consumption distribution.

Political affiliation also correlated with compliance. Households in counties that supported the Governor the least in the 2018 election responded the least to the request, and households in counties that supported the Governor the most had high levels of compliance.

Thus, it appears that political polarization can lead to defiance from outsider groups; however, this does not outweigh the benefits of the Governor's amplification of the emergency appeal via social media and the text alert system.

Ultimately, the 2019 Michigan polar vortex crisis was resolved by a combination of residential and non-residential demand reductions and supply side efforts. After the crisis, the Governor of Michigan issued an executive order transferring energy emergency response management from the Michigan Energy Agency to the Michigan Public Service Commission (MPSC) and commissioned an assessment of Michigan energy resources from the MPSC (MPSC, 2019a). The report includes an overview of energy supply systems for natural gas, electricity, and propane, as well as a section on energy emergency management. The section on emergency management states in general terms that utilities can pursue a variety of curtailment strategies to reduce demand, including voluntary requests and rate increases but the guidelines are vague.

This paper shows that emergency demand response programs can help provide stability during times of crisis, but the efficacy of the program depends heavily on its design and implementation. While the emergency request in Michigan was largely successful and the worst-possible outcome was averted, the low overall level of compliance and perverse reference-point effects highlight the need for well-designed emergency measures to reduce energy demand. To be successful, a voluntary emergency demand-response program needs a communication platform that enables it to reach households, can induce compliance from households that receive the request, and that has a demonstrated effectiveness so that utilities and balancing authorities know the size of the demand reductions to expect. Rather than relying on voluntary requests with unknown efficacy, utilities should develop, test, and optimize programs that can be called upon when needed.

To avoid compliance issues, utilities can invest in voluntary programs that eliminate compliance barriers by purchasing centralized control of energy consumption long before emergency events occur. Interruptible-load demand response programs are not new, but applications involving smart thermostats may provide a new opportunity to enhance emergency management. For households who do not wish to surrender control or who have conventional thermostats, incentive-based emergency curtailment program design remains critical.

We see several results as likely to generalize beyond energy consumption. First, compliance with requests is likely to be higher when the messenger is a trusted public figure. This effect may be reduced by political polarization or distrust of institutions. Second, a low-cost method of widespread emergency notification such as the cell phone alert system is key for communicating timely requests during a crisis. Emergency communication via social media is likely to suffer from low reach and must compete with other content for visibility. The incentives for compliance also matter. Requests for uniform compliance goals are likely to be too ambitious for some and too conservative for others. Instead, a simple request for a marginal contribution to the public good or a menu of marginal contributions that can be dialed up or down avoids this problem without the need to explicitly tailor requests. Finally, this event highlights the need for testing emergency planning and incorporating design elements that explicitly consider economic incentives and behavioral responses.

References

Abadie, A., S. Athey, G. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? NBER working paper 24003, National Bureau of Economic Research.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics 95*(9), 1082–1095. Special Issue: The Role of Firms in Tax Systems.

Allcott, H., L. Boxell, J. Conway, M. Gentzkow, M. Thaler, and D. Yang (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. NBER working paper 26946, National Bureau of Economic Research.

Allcott, H. and T. Rogers (2014, October). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review 104*(10), 3003–37.

Allcott, H. and D. Taubinsky (2015, August). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review 105*(8), 2501–38.

Anderson, S. T. and J. M. Sallee (2011, June). Using loopholes to reveal the marginal cost of regulation: The case of fuel-economy standards. *American Economic Review 101*(4), 1375–1409.

Barrios, J. and Y. Hochberg (2020). Risk perception through the lens of politics in the time of the covid-19 pandemic. NBER working paper 27008, National Bureau of Economic Research.

Beatty, T. K. M., J. P. Shimshack, and R. J. Volpe (2019). Disaster preparedness and disaster response: Evidence from sales of emergency supplies before and after hurricanes. *Journal of the Association of Environmental and Resource Economists 6*(4), 633–668.

Blackport, R. and J. Screen (2020). Weakened evidence for mid-latitude impacts of arctic warming. *Nature Climate Change 10*, 1065–1066.

Blonz, J., K. Palmer, C. Wichman, and D. Wietelman (2021). Smart thermostats, automation, and time-varying prices. Technical Report 21-20, Resources for the Future.

Brandon, A., C. M. Clapp, J. A. List, R. Metcalfe, and M. Price (2021). Smart tech, dumb humans: The perils of scaling household technologies. Working paper.

Brent, D. A., J. H. Cook, and S. Olsen (2015). Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials. *Journal of the Association of Environmental and Resource Economists 2*(4), 597–627.

Breza, E., F. C. Stanford, M. Alsan, B. Alsan, A. Banerjee, A. G. Chandrasekhar, S. Eichmeyer, T. Glushko, P. Goldsmith-Pinkham, K. Holland, E. Hoppe, M. Karnani, S. Liegl,

T. Loisel, L. Ogbu-Nwobodo, B. Olken, C. Torres, P. L. Vautrey, E. T. Warner, S. Wootton, and E. Duflo (2021). Effects of a large-scale social media advertising campaign on holiday travel and covid-19 infections: a cluster randomized controlled trial. *Nature Medicine 27*, 1622 – 1628.

Brown, Z., N. Johnstone, I. Haščič, L. Vong, and F. Barascud (2013). Testing the effect of defaults on the thermostat settings of oecd employees. *Energy Economics 39*, 128–134.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997, 05). Labor Supply of New York City Cabdrivers: One Day at a Time*. *The Quarterly Journal of Economics 112*(2), 407–441.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008, 08). Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics 90*(3), 414–427.

Carlsson, F., C. Gravert, O. Johansson-Stenman, and V. Kurz (2021). The use of green nudges as an environmental policy instrument. *Review of Environmental Economics and Policy 15*(2), 216–237.

Cialdini, R. (2006). *Influence: The Psychology of Persuasion, Revised Edition.* Harper Business.

Consumers Energy Company (2019). Ray compressor station fire, Jan. 30, 2019. Technical report. Report submitted to Michigan Public Service Commission Case No. U-20463.

CQ Press (2014-2019). Voting and elections collection. Governor election returns, county detail by year, http://library.cqpress.com/elections/download-data.php, accessed 05-2020.

DesOrmeau, T. (2019). Michiganders answered call, cut gas usage 10 percent after emergency plea. Technical report, MLive.

Edwards, J. T. and J. A. List (2014). Toward an understanding of why suggestions work in charitable fundraising: Theory and evidence from a natural field experiment. *Journal of Public Economics 114*, 1–13.

EIA (2019a). Extreme cold in the Midwest led to high power demand and record natural gas demand. Technical report, United States Energy Information Administration.

EIA (2019b). Record cold temperatures in the upper midwest increase u.s. natural gas heating demand. Technical report, United States Energy Information Administration.

Engström, P., K. Nordblom, H. Ohlsson, and A. Persson (2015, November). Tax compliance and loss aversion. *American Economic Journal: Economic Policy 7*(4), 132–64.

Farber, H. S. (2008, June). Reference-dependent preferences and labor supply: The case of new york city taxi drivers. *American Economic Review 98*(3), 1069–82.

Ferraro, P. J., J. J. Miranda, and M. K. Price (2011, May). The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. *American Economic Review 101*(3), 318–22.

Ferraro, P. J. and M. K. Price (2013, 03). Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment. *The Review of Economics and Statistics 95*(1), 64–73.

Foster, A., E. Gutierrez, and N. Kumar (2009, May). Voluntary compliance, pollution levels, and infant mortality in mexico. *American Economic Review 99*(2), 191–97.

Foster, A. D. and E. Gutierrez (2013, May). The informational role of voluntary certification: Evidence from the mexican clean industry program. *American Economic Review 103*(3), 303–08.

Ge, Q. and B. Ho (2019). Energy use and temperature habituation: Evidence from high frequency thermostat usage data. *Economic Inquiry 57*(2), 1196–1214.

Gray, K. (2019). How the Consumers Energy polar vortex emergency unfolded. Technical report, The Detroit Free Press.

Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics 148*, 14–31.

Harding, M. and A. Hsiaw (2014). Goal setting and energy conservation. *Journal of Economic Behavior & Organization 107*, 209–227.

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

Iyengar, S., Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science 22*(1), 129–146.

Kim, J. and S. S. Oh (2015). Confidence, knowledge, and compliance with emergency evacuation. *Journal of Risk Research 18*(1), 111–126.

Kotchen, M. J., G. Jacobsen, and M. Vandenbergh (2012). The behavioral response to voluntary provision of an environmental public good: Evidence from residential electricity demand. *European Economic Review 56*(5), 946–960.

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

Luyben, P. D. (1982). Prompting thermostat setting behavior: Public response to a presidential appeal for conservation. *Environment and Behavior 14*(1), 113–128.

Meier, A., U. Tsuyoshi, M. Pritoni, L. Rainer, A. Daken, and D. Baldewicz (2019). What can connected thermostats tell us about american heating and cooling habits? In *ECEEE Summer Study Proceedings*. European Council for an Energy Efficient Economy.

Montero, J. (1999). Voluntary compliance with market-based environmental policy: Evidence from the U.S. acid rain program. *Journal of Political Economy 107*(5), 998–1033.

MPSC (2019a). Michigan statewide energy assessment. Technical report, Michigan Public Service Commission.

MPSC (2019b). Michigan's statewide energy assessment fact sheet. Technical report, Michigan Public Service Commission.

NOAA (2019). The science behind the polar vortex: You might want to put on a sweater. Technical report, United States Department of Commerce National Oceanic and Atmospheric Administration.

NOAA (2021). U.s. billion-dollar weather and climate disasters. https://www.ncdc.noaa.gov/billions/ accessed 09-06-2021).

Perryman Group (2021). Preliminary estimates of economic costs of the february 2021 Texas winter storm. Technical report, Perryman Group. https://www.perrymangroup.com/media/uploads/brief/perryman-preliminary-estimates-of-economic-costs-of-the-february-2021-texas-winter-st pdf accessed 05-24-2021.

Seibold, A. (2021, April). Reference points for retirement behavior: Evidence from german pension discontinuities. *American Economic Review 111*(4), 1126–65.

Sigman, H. and H. F. Chang (2011, May). The effect of allowing pollution offsets with imperfect enforcement. *American Economic Review 101*(3), 268–72.

Thakral, N. and L. T. Tô (2021, August). Daily labor supply and adaptive reference points. *American Economic Review 111*(8), 2417–43.

Thaler, R. and C. Sunstein (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *Science 211*(4481), 453 – 458.

US Census Bureau (2019). American community survey 1-year sample.

Whitehead, J. C., B. Edwards, M. V. Willigen, J. R. Maiolo, K. Wilson, and K. T. Smith (2000). Heading for higher ground: factors affecting real and hypothetical hurricane evacuation behavior. *Global Environmental Change Part B: Environmental Hazards 2*(4), 133–142.

Wichman, C. J., L. O. Taylor, and R. H. von Haefen (2016). Conservation policies: Who responds to price and who responds to prescription? *Journal of Environmental Economics and Management 79*, 114–134.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics 141*(2), 1281–1301.

Zou, E. Y. (2021, July). Unwatched pollution: The effect of intermittent monitoring on air quality. *American Economic Review 111*(7), 2101–26.

Appendices

A Robustness checks

In this section, we test the sensitivity of the average treatment effect estimates to specification, sample selection, potential spillovers, and membership in Ecobee's "eco+" energy efficiency program. We find that the estimates are not affected by these potential confounders.

First, we examine the effect of specification choice on the difference-in-difference estimates. In the main text, we display results using a two-way fixed effects approach. Tables 4

Table 4: Alternate difference-in-differences specifications with temperature as the outcome variable.

	Thermostat setting DID			
	(1)	(2)	(3)	(4)
VARIABLES	Model 1	Model 2	Model 3	Model 4
Michigan	-0.493**	-0.497**		
	(0.097)	(0.097)		
Post	0.165**	0.419**	0.325**	0.487**
	(0.030)	(0.044)	(0.024)	(0.023)
Michigan x Post	-0.826**	-0.825**	-0.834**	-0.853**
	(0.049)	(0.050)	(0.049)	(0.047)
Constant	67.380**	67.732**	67.668**	67.117**
	(0.061)	(0.103)	(0.025)	(0.026)
Observations	7,604,357	7,604,174	7,604,174	7,604,174
R-squared	0.004	0.006	0.671	0.694
Weather		YES	YES	YES
Household FE			YES	YES
Day of week				YES
Hour of day				YES

Robust standard errors clustered at the city level.
** $p<0.01$, * $p<0.05$

Table 5: Alternate difference-in-differences specifications with compliance as the outcome variable.

	Compliance LPM DID			
	(1)	(2)	(3)	(4)
VARIABLES	Model 1	Model 2	Model 3	Model 4
Michigan	0.042**	0.042**		
	(0.008)	(0.008)		
Post	0.005	-0.025**	-0.019**	-0.038**
	(0.003)	(0.004)	(0.002)	(0.002)
Michigan x Post	0.111**	0.112**	0.112**	0.114**
	(0.006)	(0.006)	(0.006)	(0.006)
Constant	0.194**	0.164**	0.162**	0.225**
	(0.004)	(0.008)	(0.003)	(0.003)
Observations	7,604,357	7,604,174	7,604,174	7,604,174
R-squared	0.003	0.005	0.432	0.459
Weather		YES	Household YES	Household YES
FE			YES	YES
Day of week				YES
Hour of day				YES

Robust standard errors clustered at the city level.
** p<0.01, * p<0.05

Table 6: Alternate difference-in-differences specifications with fan run time as the outcome variable.

	Fan minutes running DID			
	(1)	(2)	(3)	(4)
VARIABLES	Model 1	Model 2	Model 3	Model 4
Michigan	-2.445**	-2.709**		
	(0.503)	(0.528)		
Post	10.547**	2.478**	1.760**	1.742**
	(0.153)	(0.139)	(0.074)	(0.076)
Michigan x Post	-2.376**	-1.698**	-1.676**	-1.665**
	(0.385)	(0.305)	(0.293)	(0.292)
Constant	20.942**	28.067**	27.893**	28.016**
	(0.397)	(0.664)	(0.122)	(0.157)
Observations	7,604,373	7,604,190	7,604,190	7,604,190
R-squared	0.017	0.051	0.600	0.617
Weather		YES	YES	YES
Household FE			YES	YES
Day of week				YES
Hour of day				YES

Robust standard errors clustered at the city level.
** $p<0.01$, * $p<0.05$

- 6 display alternative specification coefficient estimates for each outcome variable. Column 1 displays estimates using a standard difference in differences specification with indicator variables for being in Michigan, being in the post-treatment period, and the interaction of these two. Column 2 adds time-varying controls for temperature and humidity, column three replaces the treatment group indicator with household fixed effects, and column four replaces the post-treatment indicator with day-of-week and hour of day indicator variables. The results do not differ substantially across all specifications for each outcome variable.

Table 7: Estimates of the regressions from equation 1 on a balanced panel of households.

Robustness check: Estimated on balanced panel

VARIABLES	(1) Thermostat setting	(2) Compliance	(3) Fan run time
Michigan x Post	-0.87**	0.12**	-1.85**
	(0.05)	(0.01)	(0.31)
Constant	67.10**	0.22**	25.26**
	(0.03)	(0.00)	(0.15)
Observations	7,117,640	7,117,640	7,117,664
R-squared	0.70	0.47	0.63
Weather	YES	YES	YES
FE	YES	YES	YES
Hour FE	YES	YES	YES

Robust standard errors clustered at the city level.
** p<0.01, * p<0.05

Next, we consider the possibility of sample selection. In the sample, 6.9% of households enter late or leave early. This is best thought of as a sample selection problem as we do not observe the households before or after these points. Because the polar vortex event was unanticipated, the missing observations are "missing completely at random" and are therefore unrelated to the error term (Wooldridge, 2007). Nonetheless, we drop households that enter the sample late, leave early, or otherwise are missing observations and estimate the average treatment effect using a balanced panel of households using the two-way fixed effects specification of equation 1. Table 7 displays the results of this estimation. As expected, the results do not differ substantially from the estimates in the main text.

Table 8: Estimates of the regressions from equation 4, allowing the treatment to potentially spill over into border counties.

	Spillovers		
	(1)	(2)	(3)
VARIABLES	Thermostat setting	Compliance	Fan run time
Michigan x Post	-0.86**	0.11**	-1.94**
	(0.05)	(0.01)	(0.32)
Border county x Post	-0.04	0.01	0.67
	(0.12)	(0.01)	(0.51)
Constant	67.10**	0.22**	25.22**
	(0.03)	(0.00)	(0.15)
Observations	7,604,174	7,604,174	7,604,190
R-squared	0.70	0.46	0.63
Weather	YES	YES	YES
Household FE	YES	YES	YES
Time FE	YES	YES	YES

Robust standard errors clustered at the city level.
** p<0.01, * p<0.05

Our next robustness check allows for the possibility of spillovers to counties bordering Michigan. Because the text alerts go to cellphones based upon the closest cell phone tower, it is possible that households living near the border in Indiana and Ohio were also treated. Illinois and Wisconsin do not border Michigan's lower peninsula. 3.4 percent of Ohio and 11.6 percent of Indiana sample households live in counties bordering Michigan. We estimate the following regression, which allows for a spillover treatment effect for households living in border counties:

$$Y_{i,t} = \alpha_i + \lambda_t + \beta D_{i,t} + \sigma S_{i,t} \gamma X_{i,t} + \varepsilon_{i,t}, \tag{4}$$

where $S_{i,t}$ is a treatment variable equal to one for households in counties that border Michigan's lower peninsula during the post-treatment period. The estimated coefficient σ on $S_{i,t}$ should be equal to zero if there are no spillovers into the bordering counties. Table 8 displays the estimated coefficients using all three outcome variables. In each regression, the estimated spillover coefficient is small and the confidence interval contains zero. The coefficient on the

Table 9: Estimates of the regressions from equation 1 omitting households enrolled in Ecobees eco+ program.

Robustness check: Estimated dropping eco-slider households

	(1)	(2)	(3)
VARIABLES	Thermostat setting	Compliance	Fan run time
Michigan x Post	-0.85**	0.11**	-1.98**
	(0.05)	(0.01)	(0.32)
Constant	67.10**	0.22**	25.33**
	(0.03)	(0.00)	(0.15)
Observations	6,992,689	6,992,689	6,992,699
R-squared	0.70	0.46	0.63
Weather	YES	YES	YES
FE	YES	YES	YES
Hour FE	YES	YES	YES

Robust standard errors clustered at the city level.
** p<0.01, * p<0.05

main treatment variable is not substantially different from the estimates in the main text. We have experimented with modeling potential spillovers as far as two counties away from the border and we do not see much difference in the estimated coefficients on the main treatment variable, so we do not display the results here. In addition to these empirical results, searches of the archives of Toledo, Ohio's main newspaper, *The Blade*, using the keywords "polar vortex" and "natural gas" does not turn up any news coverage of the Michigan event. These results suggest that any potential spillover effects are not affecting the estimates.

Finally, we analyze whether the results change if we omit households enrolled in Ecobees eco+ program (8.0% of households). The eco+ program manipulates the thermostat setting based upon households' schedules to reduce energy bills. We estimate the average treatment effect on households that do not belong to the eco+ program using the two-way fixed effects specification of equation 1. Table 9 displays the results of this estimation. Again, the results to do not differ substantially from the estimates in the main text.

B Placebo analyses

Ten days before the polar vortex, Michigan experienced a similar cold wave that did not coincide with a supply-side emergency causing an emergency request for voluntary thermostat reductions. Figure 2a displays mean daily temperatures for sample households in Michigan in January 2019. Temperatures on January 20-21 dropped from 25°F to below 10°F, making these days a perfect placebo event for the January 30-31 emergency. Because there was no emergency request to reduce thermostat settings on the 20th and 21st, we would expect to see no difference in heating behavior for Michigan and control states.

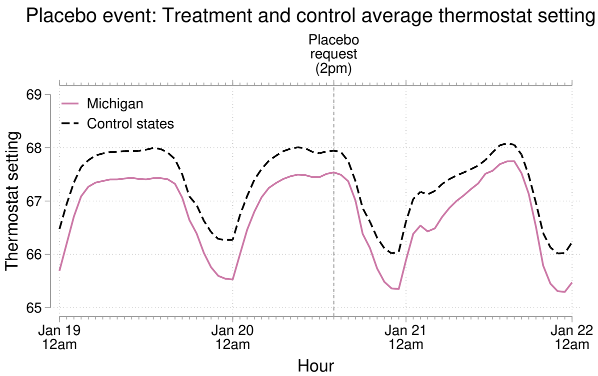
B.1 Average treatment effects

Figure 8 displays average thermostat settings, fraction of households with thermostat settings below 65 degrees F, and fan running times for Michigan and control states from January 19-22. Unlike in the main text, we do not see a change in heating behavior between Michigan and the controls states after the placebo treatment time.

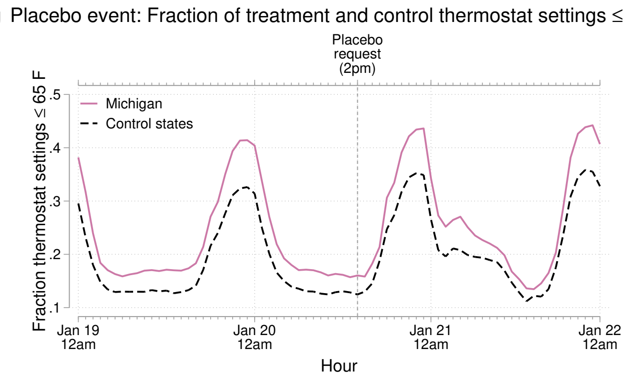
We then estimate the two-way fixed-effects specification of regression equation 1 for the placebo event. To do so, we include all observations observed between January 1st and January 21st, 2019 and treat 2:00 pm on January 20th as the placebo treatment time. Table 10 displays the results of this estimation using thermostat setting, an indicator variable for having the thermostat below 65 degrees F, and fan running time as outcome variables. As expected, the estimates are close to zero. This placebo procedure demonstrates that households in Michigan respond similarly to cold spells as households in the control states absent a request to reduce energy consumption.

B.2 Five-minute thermostat settings

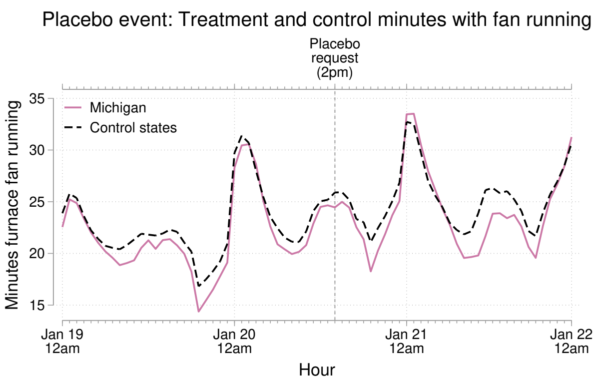
In figure 9a, we plot five-minute thermostat setting data for Michigan and the control states between 12:00 pm on January 20th and 11:59 pm on January 21st. In figure 9b, we plot a difference-in-differences estimate of the placebo treatment effect, which we construct



(a)



(b)



(c)

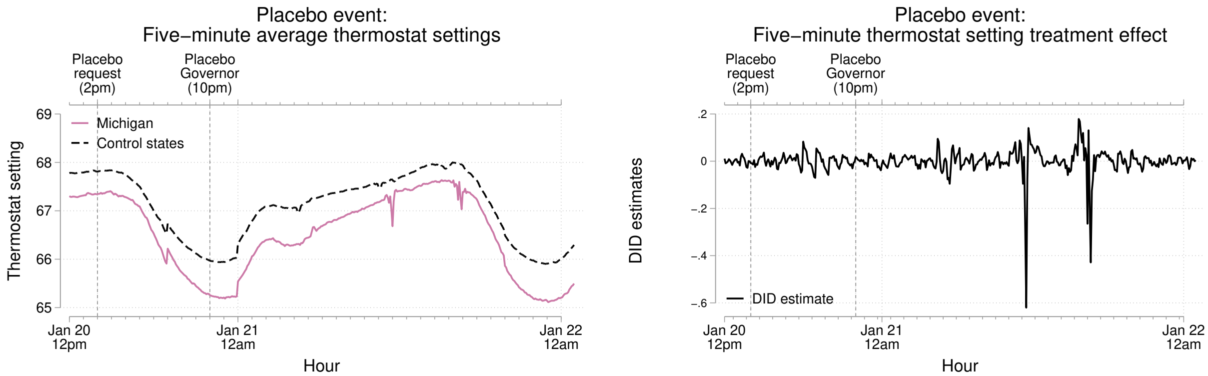
Figure 8: Sample average values of the outcome variables for treatment and control households, January 19 - 22. Panel (a) plots average thermostat settings, panel (b) plots the fraction of households with thermostat settings below 65°F, and panel (c) plots the average furnace fan run time in each hour. The vertical dashed line indicates 2:00 pm, the beginning of the placebo event.

Table 10: Estimates of the regressions from equation 1 on the sample of households observed between January 1st and January 21st, treating January 20th at 2:00 pm as the placebo treatment time. We expect the estimates from this placebo estimation to be close to zero.

Placebo estimates from Jan 20-21 cold wave

VARIABLES	(1) Thermostat setting	(2) Compliance	(3) Fan run time
Michigan x Post	0.05	0.00	0.21
	(0.04)	(0.00)	(0.21)
Constant	67.00**	0.23**	23.94**
	(0.04)	(0.00)	(0.16)
Observations	5,125,897	5,125,897	5,125,922
R-squared	0.72	0.48	0.64
Weather	YES	YES	YES
FE	YES	YES	YES
Hour FE	YES	YES	YES

Robust standard errors in parentheses
** p<0.01, * p<0.05



(a)

(b)

Figure 9: (a): Five-minute sample mean thermostat settings for Michigan and control households from 12:00 pm January 20th - 11:59 pm January 21st. (b): Five-minute difference-in-differences estimate.

as the difference between five-minute thermostat setting for Michigan and control states during the event minus the average difference in same hour-of-day and day-of-week five-minute thermostat settings before the event. In these figures, we see a discrete increase in thermostat setting in the five-minute periods beginning at 7:00 pm and 12:00 am for both treatment and control households. These discrete changes correspond to commonly programmed times for the thermostat to automatically change. Other than in a few five-minute periods on January 21st, the measured treatment effect is zero in the placebo period. These treatment effects are driven by seemingly spurious five-minute jumps in the average thermostat setting for Michigan households. Overall, these placebo plots demonstrate the validity of the difference-in-differences approach and verify that discrete increases at 7:00 pm and 12:00 am are common and not artifacts of the emergency event.

B.3 Heterogeneity analysis

One concern that we had was that the estimates from the heterogeneity analysis in the main text reflect statistical reversion to the mean rather than a meaningful pattern of results for the sub-groups (particularly for the baseline thermostat-setting results). To test this, we estimate the heterogeneity regression analysis from equation 3 during the placebo period.

Table 11 displays the estimated coefficients and bootstrapped standard errors that account for the first-stage estimation of the baseline thermostat setting. The magnitude of the estimated coefficients in the placebo analysis are almost all two to five times smaller than during the polar vortex and do not display as clear trends as you move away from 65°F. Thus, we conclude that mean reversion may play a small role in the main heterogeneity estimates but the effect is not large enough to alter our conclusions in section 4.3.

Table 11: Results from the estimation of equation 3 using the placebo cold wave.

	(1) Thermostat setting	(2) Compliance	(3) Fan run time
40-45% Democrat X Treatment	-0.12	0.014	0.47
	(0.14)	(0.013)	(0.60)
45-50% Democrat X Treatment	-0.41*	0.024	-0.023
	(0.18)	(0.017)	(0.82)
50-55% Democrat X Treatment	-0.37**	0.027*	0.91
	(0.13)	(0.014)	(0.68)
55-60% Democrat X Treatment	-0.13	0.026	1.71
	(0.20)	(0.021)	(0.94)
60-65% Democrat X Treatment	-0.29	0.017	2.20
	(0.20)	(0.021)	(1.24)
65-70% Democrat X Treatment	-0.48	0.0058	2.21
	(0.29)	(0.027)	(1.48)
70-75% Democrat X Treatment	-0.32	0.052	1.39
	(0.30)	(0.030)	(1.20)
59 F or lower expected X Treatment	1.33**	-0.0060	0.82
	(0.30)	(0.013)	(0.51)
59-61 F expected X Treatment	0.63**	-0.062**	0.33
	(0.15)	(0.0097)	(0.48)
61-63 F expected X Treatment	0.39**	-0.048**	0.69
	(0.089)	(0.0094)	(0.41)
65-67 F expected X Treatment	-0.21**	0.17**	-1.03**
	(0.049)	(0.0099)	(0.32)
67-69 F expected X Treatment	-0.29**	0.10**	-0.95**
	(0.053)	(0.0095)	(0.33)
69-71 F expected X Treatment	-0.30**	0.078**	-0.60
	(0.055)	(0.0092)	(0.41)
71-73 F expected X Treatment	-0.35**	0.065**	-0.95*
	(0.070)	(0.0095)	(0.48)
73-75 F expected X Treatment	-0.69**	0.055**	-0.79
	(0.12)	(0.010)	(0.97)
Higher than 75 F expected X Treatment	-0.096	0.055**	-1.90
	(0.24)	(0.011)	(1.76)
Observations	4,751,555	4,751,555	4,751,585
R-squared	0.81	0.66	0.64
FE	YES	YES	YES
Hour	YES	YES	YES
Controls	YES	YES	YES
Expected thermostat level	YES	YES	YES

Standard errors cluster-bootstrapped at the city level.
** p<0.01, * p<0.05