Coordinating Energy Efficiency and Incentive-based Demand Response

Eric M. Webb[1], Owen Q. Wu[2], Kyle D. Cattani[2]

[1] Carl H. Lindner College of Business, University of Cincinnati, Cincinnati, OH 45221, USA

[2] Kelley School of Business, Indiana University, Bloomington, IN 47405, USA

eric.webb@uc.edu; owenwu@indiana.edu; kcattani@indiana.edu

August 2, 2019

Problem definition: Energy efficiency programs and demand response programs, two popular approaches to energy demand management, are traditionally designed and evaluated independently. Breaking with tradition, we study the interactions between energy efficiency upgrades and participation in incentive-based demand response programs. We re-examine the energy efficiency gap in light of demand response participation. We also illustrate how to coordinate the programs to maximize their combined benefits.

Academic/practical relevance: Billions of dollars are spent annually in the U.S. on both energy efficiency incentives and demand response payments. Without coordination, the incentives from one program could undercut the effectiveness of another program. We examine ways to coordinate the incentives.

Methodology: We use a sequential optimization model which accounts for the difference in time scale between long-term energy efficiency investments and daily demand response participation and minimizes discounted costs to an industrial firm as well as to society.

Results: A larger demand response incentive leads to a reduced investment in energy efficiency. Even if the energy efficiency gap is closed by taxes or subsidies, the firm's impact on societal costs is not minimized unless the demand response incentive is properly set. Both energy efficiency and demand response programs contribute significantly to the reduction of societal costs.

Managerial implications: Policies aiming to close or reduce the energy efficiency gap, such as investment subsidies and carbon taxes, may fail to achieve their desired outcomes when firms also participate in demand response. We provide theoretical support for jointly determining energy efficiency and demand response incentives, in both unconstrained and constrained policy situations, to maximize the combined benefits of both demand management programs.

Key words: Energy efficiency, demand response, energy efficiency gap, operational flexibility

1. Introduction

Energy efficiency (EE) programs and demand response (DR) programs are designed to modify patterns of electrical load, including the timing and level of electricity demand (Energy Information Administration 2017). Modifying load patterns can lessen the strain on the electricity grid, improve grid reliability, and reduce overall power generation costs. EE and DR programs typically are designed and implemented independently, despite their joint effect on load patterns. Without understanding the interactions between these programs, efforts to promote and implement them may fail to achieve desired outcomes. In this paper, we examine the potential conflicts between EE and DR programs, revisit the "energy efficiency gap" in conjunction with DR participation, and show the benefit of jointly designing policies that promote EE and DR.

EE is a way of using less energy to provide the same service. Examples include using more efficient bulbs to provide the same amount of light or upgrading an alloy smelter's furnaces to reduce the electricity consumption per ton of alloy produced. EE benefits society by decreasing demand for electricity, thereby reducing the environmental externalities of electricity generation.

For a variety of reasons, firms install less EE than the amount that would be considered optimal for society (Jaffe and Stavins 1994, Gerarden et al. 2017). The difference between the socially desirable level of EE investment and the level chosen by the firm is referred to as the EE gap. In recent years, enormous incentives have been created for EE investments, in part to close the EE gap. The American Recovery and Reinvestment Act of 2009 allocated about $20 billion to EE programs (White House Council of Economic Advisors 2016). Days after the 2015 Climate Change Conference in Paris, the U.S. government rolled out the "largest energy-saving standard in history," targeting commercial air conditioners and furnaces (Mooney 2015). This paper shows that the EE gap needs to be reevaluated in the presence of DR.

DR programs are used to modify electrical load patterns. In particular, DR programs are designed to induce reductions in electricity demand during times of electric system contingencies and/or high demand, thereby reducing the need to call upon less efficient peaking power plants with high marginal costs. In this paper, we focus on voluntary, incentive-based DR programs, such as the Real-Time Price Response program at NextEra Energy Services (2017) and the PowerShare QuoteOption program at Duke Energy (2017). In these programs, when an extremely high load is forecasted, DR-participating firms are notified in advance to reduce load. If a firm reduces its load, it is paid for its curtailment at an announced rate, which may depend on the wholesale market price. Other types of DR programs are reviewed in §2, and an extension to mandatory DR programs is discussed in §7.

EE investments reduce load at all times, while DR programs focus on reducing load during the most expensive times without necessarily reducing the total energy consumed (i.e., electrical load may be shifted from peak to off-peak times). Both EE and DR reduce peak demand and total energy generation costs. The combined economic benefits of EE and DR programs exceed what would be possible with either program alone (Smith and Managan 2012, Nadel 2017). Thus, both demand-side management programs have become increasingly popular. An annual electric power industry survey (Energy Information Administration 2016) shows that from 2013 to 2015 in the U.S., EE programs for industrial firms achieved energy savings of about 3 million megawatt-hours (MWh) annually from new EE programs or by new participants in existing EE programs. The DR potential of industrial firms increased from 14.8 gigawatts (GW) in 2013 to 17.2 GW in 2015.

Unfortunately, EE and DR programs are typically designed and implemented separately (Vine 2008), and research on the structural interactions between EE and DR is very limited. Significant efforts have been undertaken to engage customers to participate in both programs (see §2 for a review). However, even if customers are fully engaged in both programs, the separation of program design remains a barrier to maximizing the combined benefits of EE and DR. This interaction between EE and DR was recently brought to the attention of the U.S. Congress. Hledik et al. (2016) describe the case of electric resistance water heaters, which were to be phased out under the 2010 EE standard set by the U.S. Department of Energy. These water heaters are often used as a DR resource, as they can be curtailed during times of peak demand if water has been pre-heated to an acceptable level. Each water heater effectively acts as a battery for load shifting, peak shaving, or to integrate more renewables onto the grid. Recognizing these benefits, the Energy Efficiency Improvement Act of 2015 made an exception for many electric resistance water heaters that are grid-enabled to provide DR.

We explore the economic interactions between incentive-based DR programs and EE investments at industrial facilities. Compared to households, industrial firms have higher energy costs as a percentage of total costs. For manufacturing industries, the cost of energy as a percentage of total production cost often exceeds 10% in many basic materials industries and reaches almost 40% in cement manufacturing (Natural Resources Canada 2006). As such, industrial firms are likely to provide the necessary effort to optimize EE investments and DR participation. For example, Olsen et al. (2010) discusses the cement manufacturing industry, in which EE upgrades can offer significant cost savings and DR is practical at interruptible manufacturing steps, such as quarrying, raw mix grinding, fuel grinding, and clinker grinding. For refrigerated warehouses in the food industry, Scott et al. (2015) advocates for concurrent consideration of EE and DR, as the magnitude of DR potential is reduced when refrigeration systems are more energy efficient.

How industrial firms respond to DR events depends on the types of processes and energy usage involved, but a common practice is to pause part of the processes and reschedule the work and energy use to a different time. For example, the authors are familiar with a large manufacturing firm in Kentucky that normally runs 14 molding machines, consuming a total of 2.3 MW; some or all of these machines can be temporarily shut down during a DR event called by Kentucky Utilities. After each DR event, the company receives an incentive payment from Kentucky Utilities and makes up for the lost production during overtime. The trend in EE in injection molding is moving away from hydraulic machines and replacing them with all-electric machines, which can reduce energy usage by 30% to 60%. However, due to the DR incentive payments, which were said to be lucrative, the company preferred to extend the use of the current fleet of machines.

To investigate the economic interactions between EE and incentive-based DR, we create a firm-level sequential optimization model, with long-term EE first being installed and daily DR participation subsequently occurring over time. We solve this model from the perspective of both a cost-minimizing firm and a societal decision maker who incorporates the environmental externalities of electricity generation into her decisions.

We summarize our theoretical and practical insights as follows. EE installations alter subsequent DR participation. Typically, when a firm becomes more efficient, it participates less in DR. This is because firms get paid based on the amount of load they curtail in a DR event relative to their typical baseline load. Becoming more efficient lowers the baseline, making DR participation less lucrative. However, an EE upgrade that also significantly decreases the costs of shifting load to overtime (e.g., it reduces the labor requirements) could lead to an alternate relationship in which the firm has more DR participation as more EE is installed. Thus, the labor-saving aspects of the new machinery installed during the EE upgrade can modify the effect of EE on DR.

However, regardless of the labor-saving aspects of EE, a higher available DR incentive leads to less upfront EE investment at the firm. Accordingly, DR programs with high incentive payments could be undermining the efficacy of EE programs that seek to close the EE gap. A societal decision maker typically desires more EE than the firm is willing to install due to the inclusion of environmental externalities in her objective function. This EE gap grows with the DR incentive as a firm that participates in DR becomes less and less willing to install EE and lower its baseline load as DR becomes more lucrative.

To maximize societal benefits from EE and DR, incentives in each program must be coordinated. EE investment subsidies/rebates and carbon taxes are two frequently discussed EE incentives to help close the EE gap. These EE incentives will be less effective for a firm that participates in DR. Further, closing the EE gap is a problematic objective when the firm participates in DR, as closing the EE gap could actually increase overall societal costs under certain circumstances. Specifically, if the EE gap is closed by decreasing the DR incentive, the lost DR participation at the firm may be more costly to society than the savings from increased EE. Only by coordinating these incentives can an appropriate amount of both EE and DR be induced.

It is often impractical to induce the firm's EE investment and DR participation to match the decisions that would be optimal for society, because EE investment subsidies, carbon taxes, and DR incentive payments involve other economic and/or political constraints. As such, we also demonstrate a constrained policy optimization. For a given, acceptable EE subsidy (or carbon tax), we show the DR incentive level that will minimize the impact of the firm's actions on societal costs.

2. Literature Review

We discuss the literature surrounding the EE gap, DR program types, and the interactions between EE and DR.

The notion of an EE gap and its causes have been discussed for decades. An EE gap is created when EE technologies that would be socially efficient are not adopted. The "energy paradox," which accounts for a portion of the total EE gap, is the apparent reality that some economical EE technologies are not adopted (Gerarden et al. 2017). Explanations for the energy paradox have been provided by Hirst and Brown (1990), DeCanio (1993), Jaffe and Stavins (1994), Gillingham et al. (2009), and Gerarden et al. (2017). Several papers have discussed ways of resolving the energy paradox for industrial firms. DeCanio (1993) recommends that the government provide informational and organizational services to improve corporate decision making. Sandberg and Söderström (2003) examine the need for decision support to facilitate and improve EE investment decisions. Muthulingam et al. (2013) examine the behavioral aspects of implementing suggestions from energy audits in manufacturing firms. Aflaki et al. (2013) identify factors essential to the effective management of industrial EE projects and provide a framework for finding and implementing EE projects. In a supply chain setting, Nguyen et al. (2018) analyze how EE assessment assistance and buyers' procurement commitment can incentivize suppliers' EE investment.

Jaffe and Stavins (1994) and Gerarden et al. (2017) further point out that even if the energy paradox is not present, the EE gap may still exist because there are other reasons why the actual EE investment deviates from the socially desirable level—particularly the environmental externalities that are not fully incorporated into the price of electricity. We contribute to the literature by exploring how DR programs affect the EE gap.

There are many benefits of DR programs, as detailed by Strbac (2008), Albadi and El-Saadany (2008), Siano (2014), and O'Connell et al. (2014). The key benefit of DR comes from the fact that it is often cheaper to forgo consumption during peak times (or shift peak consumption to off-peak times) than to bring peaking power plants online. DR programs can be classified into price-based programs and incentive-based programs (U.S. Department of Energy 2006). Price-based DR programs are based on dynamic pricing of retail electricity to induce peak-load reductions (Ata et al. 2018). Prices can be preset for defined time blocks (time-of-use pricing, see Kök et al. 2016) or matched to the wholesale market price (real-time pricing, see Lohmann and Rebennack 2017) or lifted only at critical times (critical peak pricing). Incentive-based DR programs reward customers for reducing their load upon request, and can be further categorized into mandatory and voluntary curtailment programs. Mandatory curtailment programs require commitment to a pre-specified load reduction for DR (with penalties for non-compliance), while voluntary curtailment programs allow a firm to choose its level of load reduction every time a DR event is initiated.

In this paper, we focus on incentive-based DR programs with voluntary curtailment. Such programs include not only the economic DR programs mentioned in the introduction, but also emergency DR programs, such as the program at the New York Independent Service Operator (ISO) (2016). Emergency DR programs are a special case of the program we model in which DR events are called only during emergency situations, when the grid is at risk of blackout due to extreme circumstances. Our model can be readily modified to cover DR programs with mandatory curtailment, including Interruptible Load, Direct Load Control, and Load as Capacity Resource programs, as discussed in §7 and analyzed in Online Appendix C. The firm's choice between voluntary and mandatory DR contracts is studied by Daniels and Lobel (2014).

Planning and scheduling production around DR participation is, in itself, an interesting operations problem. Under a time-of-use price-based DR program, Fernandez et al. (2013) examine the build-up of buffer stock at certain machines to allow the shutdown of other machines during DR without affecting throughput. Under a voluntary, incentive-based DR program, Chao and Chen (2005) consider whether or not to shut down production during DR events, based on the inventory level, the demand process, and the DR incentive. Mohagheghi and Raji (2015) offer an optimization module that verifies whether, and by how much, an industrial plant can curtail its load while meeting production constraints.

Initial coordination efforts have been made to increase customers participation in both EE and DR programs. York and Kushler (2005) and Goldman et al. (2010) emphasize that EE and DR programs can be mutually reinforcing from the customers' perspective. For example, the information technologies used to market and monitor DR programs can also be used to help consumers understand their energy use and associated costs, thereby encouraging EE investment. Smith and Managan (2012) suggest a positive feedback loop between the programs, whereby revenue from DR participation funds EE improvements at the firm, which raise awareness of energy use and result in more DR participation. Goldman et al. (2010) explain that coordination between EE and DR can occur via combined program offerings, marketing, and education, via initiatives of private firms, and via building codes and appliance standards. In this paper, we consider a firm that is already engaged in both EE and DR, and we analyze how the DR incentive may affect the firm's EE investment decisions and how the EE investment may affect DR participation.

The literature on the structural interactions between EE and DR is very limited. King and Delurey (2005) find that EE results in peak load reductions, but cannot be dynamically controlled; DR offers dynamic control but its effect on total consumption is small on average, with notable variations across programs. York and Kushler (2005) and Smith and Managan (2012) discuss potential structural conflicts between EE and DR programs. If DR participants are paid based on the amount of load they reduce, as measured from a baseline, there can be a disincentive to take EE actions that might lower the baseline. Jewell (2014) simulates a residential DR program and finds that improved house thermal insulation reduces the home energy use and the effectiveness of the DR program, whereas improving the air conditioning efficiency may increase or decrease the effectiveness of the DR program. The above papers do not consider program design. In general, EE and DR joint program design warrants more research attention, and we complement these papers by examining the societal benefits of jointly designing EE and DR incentives to account for interactions between the two programs.

The joint design of EE and DR programs requires a change in regulatory policies. Vine (2008) calls for breaking down policy silos and changing the regulatory environment for integrating EE, DR, renewable energy, and climate change mitigation. Encouragingly, the California ISO, the California Public Utilities Commission, and the California Energy Commission have been developing a cross-agency DR and EE workplan to ensure alignment among all stakeholders (California ISO 2013).

3. Model Setup

In this section, we present a typical industrial setting and describe how EE and DR affect a firm's energy consumption profile and production schedule. Consider a manufacturing firm which must meet a production target each day. Based on the target, the firm determines a production schedule. If the normal production schedule is disrupted, overtime work (detailed in §3.2) will occur after the normal schedule. For example, the firm may schedule two eight-hour shifts per day, and overtime work can be scheduled as needed after the second shift.

We consider a planning horizon of N days. We assume the production target is the same each day, and let T denote the duration, in hours, of the normal production schedule that is adequate to meet the daily production target, e.g., $T = 16$ hours for two eight-hour shifts. Let P_0 denote the power, measured in megawatts (MW), needed for normal production. The assumptions of constant production target and constant power are not essential, but facilitate the analysis. Time-varying production targets and/or power consumption will not qualitatively alter the key tradeoffs.

Let c, measured in dollars per megawatt-hour (MWh), denote the fixed retail rate of electricity. Thus, without EE improvement, the firm's daily electricity cost is cP_0T.

3.1 Energy Efficiency (EE)

To reduce its electricity cost while maintaining the same output, the firm may install EE improvements, which require a one-time upfront investment cost and reduce future energy use. The EE installations reduce the power draw from P_0 to $P_z \stackrel{\text{def}}{=} (1-z)P_0$, where $z \in [0, 1)$ represents the proportion of power saved. We refer to z as the *EE improvement level.* For analytical convenience, we allow z to be a continuous variable, which is an approximation of the reality in which improvements can be chosen from an EE audit list (Muthulingam et al. 2013). The decision of z is made and EE improvements are installed prior to the first day of our planning horizon, and the EE savings are realized in days 1 through N.

Saving power by a proportion of z requires an upfront investment cost $I(z)$, which satisfies $I(0) = 0$ and $I(z) \to \infty$ as $z \to 1$. Furthermore, $I(z)$ is increasing and convex in z, because the firm would install improvements in order of cost-effectiveness of energy savings. We assume $I(z)$ is incurred at the beginning of day 1. If any machinery installed during the EE upgrade has a lifetime that is shorter (or longer) than N days, then a multiple (or a fraction) of the upgrade cost of the machinery will be accounted in $I(z)$.

EE investments can also provide productivity benefits, such as reduced maintenance and labor costs (Boyd and Pang 2000, Worrell et al. 2003). We include these benefits during normal work hours as a negative cost in $I(z)$, while the potential reduction in the overtime cost will be modeled in the next section.

3.2 Demand Response (DR)

During times of peak demand, when society's demand for electricity surges, the firm may be offered an incentive to curtail load. While EE improvement is a long-term decision (z is fixed over the planning horizon), DR participation is a short-term decision that may occur each day.

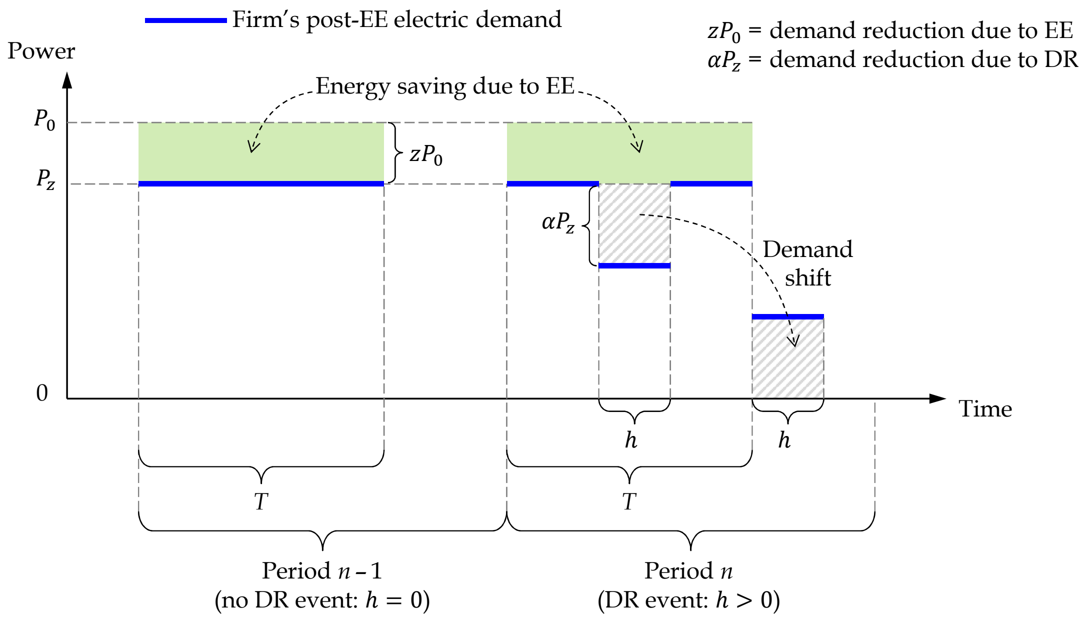
At the beginning of the day, the firm is notified whether (and if so, when) a DR event will occur. The duration of the DR event is H hours ($H = 0$ if no DR event occurs), and the firm is offered an incentive payment of R dollars per MWh of curtailment during the DR event. Both H and R are random a priori and realized at the beginning of the day, with realizations denoted as $h \in [0, h_{\max}]$ and $r \in [r_{\min}, r_{\max}]$. The incentive rate R can be predetermined by utilities or adjusted based on the wholesale market price. We assume that the event duration, H, and incentive rate, R, are independent. For analytical convenience, we assume that H and R vary over time according to a stationary process, and thus we omit time subscripts for H and R. Our analysis can be generalized to allow H and R to be correlated and follow nonstationary processes.

Given realizations of H and R (h and r), the firm decides $\alpha \in [0, 1]$, the proportion of the firm's energy use to curtail during the DR event, i.e., the firm curtails $\alpha P_z h$ MWh during the DR event. We refer to α as the firm's *DR participation level*. We assume α is a continuous decision variable. The firm curtails its power consumption during DR events by delaying production. To meet the daily production target, the firm must ensure that production curtailed during the DR event is made up through overtime. Figure 1 illustrates energy demand shift in response to a DR event.

The costs associated with the production schedule shift can be administrative (rescheduling the work), managerial (ensuring all work is still completed accurately), and compensatory (workers garner overtime pay). These costs typically scale with the amount of work shifted, which can be measured by αh (i.e., portion α of the work is shut down for h hours). Making an EE improvement to upgrade a machine may reduce the amount of labor necessary to operate the machine, which would reduce the overtime labor costs. Let $O(x, z)$ be the aforementioned costs of moving $x = \alpha h$ amount of work to overtime when the EE improvement level is z. $O(0, z) = 0$ and $O(x, z)$ is increasing in x for any given z.

In reality, overtime may also be caused by demand spikes and/or supply shortages. To capture the tradeoff between the DR incentive payment and the cost of shifting production, we assume the only reason for overtime is participation in DR.

Figure 1: Firm's power consumption under EE improvement and DR participation



For analytical convenience, we assume that $I(z)$ is differentiable and strictly convex in z, and that $O(x, z)$ is twice differentiable in (x, z) and strictly convex in x for any given z, which reflects the reality that the marginal cost of shifting additional production and workers is likely to be higher when more work has already been shifted to overtime. The derivatives are denoted as $O_1(x, z) \equiv \partial O(x, z)/\partial x$, $O_2(x, z) \equiv \partial O(x, z)/\partial z$, and $O_{12}(x, z) \equiv \partial^2 O(x, z)/\partial x \partial z$.

3.3 Power Generation Cost and Demand Response

Although the firm pays a fixed retail price for electricity, not all megawatts impose the same cost upon society. During times of peak demand, expensive peaking power plants must be used. The marginal generation cost of these peaker plants can be over \$200 per MWh, which is much higher than the marginal cost during off-peak hours—typically \$30 to \$50 per MWh (see an example in EIA 2012). Considering environmental externalities, the variable costs of peaker plants is even higher.

DR offers a demand-side alternative to running peaker plants. The duration of a DR event, defined as H in §3.2, corresponds to the duration when expensive generators are expected to be operating. Many factors, including weather and grid maintenance, influence electricity demand and electric system status. Demand forecasts and electricity day-ahead markets help utilities decide whether to announce a DR event at the beginning of a day.

To capture the essential tradeoffs, we assume that the marginal cost of power generation has two levels. Let G_p and G_b denote the marginal cost of power generation during the DR event and outside of the DR event window, respectively (subscript p for 'peak' and b for 'baseload'). To approximate the true cost of power generation to society, we also include monetized environmental externalities in G_p and G_b. Generators running on fossil fuels can lead to emissions of CO_2, SO_2, NOx, mercury, and particulates, which harm the environment and human health. The vintage of the power plants and the equipped emission controls greatly influence the emission intensities. Older peaker plants without advanced emission controls can still be called upon occasionally, leading to significantly higher environmental impacts.

Monetizing environmental externalities is challenging and prior research tends to monetize a portion of the damages and give a wide range of estimates (National Research Council 2010, Epstein et al. 2011). For our purpose, we focus on the case of $G_p > G_b$, i.e., the marginal impact of power generation is higher during the DR event. This is true in many regions, such as New England and New York, where petroleum-fueled generators are engaged during DR events and natural gas-fired generators serve as marginal units outside of the DR event windows. In our numerical examples in §5 and §6, we report results based on a set of plausible parameters: $G_b = \$200$/MWh and $G_p = \$600$/MWh. We have conducted numerical tests for a wide range of parameters and found that the qualitative insights are robust. We also consider the case of $G_p \leq G_b$, which corresponds to the situation when coal-fired generators are the dominant producers while natural gas is used during peak times.

We remark that our model is a firm-level model that reflects the decision process of a typical industrial firm. Based on our conversations with managers at a few large industrial firms, it is reasonable to assume that the frequency and length of DR events, as well as the DR incentive rate, are exogenous to the actions of individual firms.

4. Energy Efficiency Decisions Without Demand Response

We begin with a conventional cost and benefit analysis that considers EE investment in isolation of DR. This isolated analysis is appropriate for firms that lack the flexibility or opportunity to participate in DR. We consider the EE decision that minimizes costs to the firm in §4.1 and the EE decision that minimizes the firm's impact on society in §4.2. We then evaluate the EE gap in §4.3. This analysis provides a benchmark against which we may evaluate the joint EE and DR decisions in §5.

4.1 Firm's Cost Minimization

The firm strives to minimize costs over the planning horizon. The key tradeoff here is that an energy efficiency investment requires an upfront cost but decreases ongoing electricity costs. The cost-minimizing firm ignores the environmental externalities of its power consumption unless such externalities are priced into the cost of electricity.

The retail cost of electricity is c per MWh consumed. Under EE improvement level z, the daily electricity cost to the firm amounts to cP_zT. The firm chooses its EE improvement level by minimizing the sum of the upfront cost of EE improvements, $I(z)$, and the discounted future electricity bills over the planning horizon:

$$\min_{z\in[0,1)} \; I(z) + \sum_{n=1}^{N} \delta_f^n cP_zT, \tag{1}$$

where $\delta_f \in (0,1]$ is the firm's daily discount factor.

Let z^{f*} be the optimal level of EE improvement for the firm. Because $I(z)$ is convex in z and $P_z = (1-z)P_0$ is linear in z, the objective in (1) is convex in z. Thus, if $z^{f*} \in (0,1)$, it satisfies the first-order condition:

$$I'(z^{f*}) = \gamma^f cP_0T, \tag{2}$$

where $\gamma^f \equiv \sum_{n=1}^{N} \delta_f^n$.

4.2 Minimizing Firm's Impact on Societal Costs

We characterize the EE improvement level at the firm that will minimize the impact of the firm's operations on society, including environmental externalities due to the firm's energy consumption. As in §4.1, EE improvement level is chosen without DR consideration.

In the two-level cost model introduced in §3.3, the marginal cost of electricity generation (including environmental externalities) is G_p per MWh during times of peak demand and G_b per MWh in all other hours. Therefore, the EE improvement level, z, should be chosen to minimize the sum of the investment cost and the expected impact of the firm's energy use on society:

$$\min_{z\in[0,1)} \; I(z) + \mathbb{E}\left[\sum_{n=1}^{N} \delta_s^n \big(P_zHG_p + P_z(T-H)G_b\big)\right], \tag{3}$$

where the expectation is taken on H, G_p, and G_b at the time of EE investment decision, and $\delta_s \in (0,1]$ is the daily societal discount factor. Note that δ_s is often higher than the firm's discount factor, δ_f, for two reasons. First, firms generally weigh short-term costs more than long-term costs, whereas society may not significantly discount future costs. Second, society cares about the cost of emissions, which may have long-lived environmental impacts, while the firm only cares about monetary costs. Arguably, the planning horizon should also be longer from the societal point of view. However, a longer planning horizon may involve EE technology changes and multiple investments, which we leave for future research.

Because G_p and G_b are independent of H, we can rewrite the problem in (3) as

$$\min_{z\in[0,1)} \; I(z) + \gamma^s\big(P_z\bar{H}\bar{G}_p + P_z(T-\bar{H})\bar{G}_b\big) \;\equiv\; I(z) + \gamma^s\bar{c}^s P_z T, \tag{4}$$

where $\gamma^s \equiv \sum_{n=1}^{N} \delta_s^n$, $\bar{H} \equiv \mathbb{E}[H]$, $\bar{G}_p \equiv \mathbb{E}[G_p]$, $\bar{G}_b \equiv \mathbb{E}[G_b]$, and $\bar{c}^s$ is the marginal cost of electricity generation averaged over the firm's operational time T, defined as

$$\bar{c}^s \equiv \big(\bar{H}\bar{G}_p + (T-\bar{H})\bar{G}_b\big)/T. \tag{5}$$

Let z^{s*} denote the optimal EE improvement level that minimizes the firm's impact on the societal costs in the absence of DR. The convexity of the objective function in z implies that, if $z^{s*} \in (0,1)$, it satisfies the first-order condition:

$$I'(z^{s*}) = \gamma^s\bar{c}^s P_0 T. \tag{6}$$

4.3 Energy Efficiency Gap Without Demand Response

Comparing z^{f*} in (2) and z^{s*} in (6) yields insights into the EE gap, defined as $z^{s*} - z^{f*}$. The following proposition shows that this gap generally exists under mild conditions. All proofs are in Online Appendix B.

Proposition 1 *In the absence of DR, the firm would choose an EE improvement level that is lower than desired by society (i.e., EE gap $z^{s*} - z^{f*} > 0$) if the firm's discounted retail cost of electricity is lower than the societal discounted marginal cost of generating electricity, averaged over the firm's operational time (i.e. $\gamma^f c < \gamma^s\bar{c}^s$).*

Thus, at least one of the following statements must be true for society to desire more EE than the firm: (a) the full cost of electricity generation to society is not being passed on to the firm through retail electricity costs ($c < \bar{c}^s$), and (b) the firm discounts future costs more than society does ($\delta_f < \delta_s$). Both inequalities tend to be true in reality. Society experiences a larger cost than the firm because the retail cost of electricity typically does not include a full accounting of electricity generation's effects on the environment. While the exact form of $\bar{c}^s$ is specific to our model, the generalization to reality intuitively holds true: society desires more EE than the firm if the average marginal generation costs (including environmental externalities) are larger than the average electricity price paid by the firm.

5. Energy Efficiency Decisions Considering Demand Response

We now consider the EE investment problem together with participation in DR. Because the EE improvement level, z, is decided upfront, whereas the DR participation level, α, depends on the realization of event duration H and incentive rate R, we formulate the problem as a sequential optimization.

5.1 Firm's Cost Minimization

We first solve for the optimal DR participation level, α, given the EE improvement level, z, the realized length of the DR event, h, and the DR incentive rate, r. Working backward, we then solve for the optimal upfront EE improvement level, z.

5.1.1 Optimal DR Under Given EE Improvement

In practice, to determine the DR payment to the firm, a business-as-usual baseline power draw is measured by meters and verified by the administrator of the DR program. In our setting, the baseline is the firm's power consumption without DR, i.e., the baseline is P_z after EE improvement. Hence, curtailing proportion α of the power during a DR event of h hours, the firm shifts $\alpha P_z h$ MWh of energy to the off-peak and receives a payment of $r\alpha P_z h$. In practice, the baseline is calculated based on the consumption over the 5 to 10 most recent non-event workdays (e.g., PJM Interconnection 2017). Thus, after EE improvements are installed, it takes a few days for the baseline to become P_z. As EE investments represent a multi-year planning horizon, assuming that the baseline becomes P_z immediately after EE upgrades will not introduce long-term inaccuracies.

Each day, the length of the DR event, h, and the incentive rate, r, are announced. The firm then decides the proportion of energy to curtail to minimize its daily operating costs, including the DR payment as a negative cost:

$$C(z,h,r) \stackrel{\text{def}}{=} cP_zT + \min_{\alpha\in[0,1]} \Big\{O(\alpha h, z) - r\alpha P_z h\Big\}, \tag{7}$$

where cP_zT is the daily retail cost of electricity and $O(\alpha h, z)$ is the overtime cost defined in §3.2. If no DR event is called ($h = 0$), $C(z,h,r)$ collapses to cP_zT. For $h > 0$, the following lemma gives the optimal response of the firm. Recall $O_1(x,z) \equiv \partial O(x,z)/\partial x$. Treating z as a given parameter, we define the inverse function of $O_1(x,z)$ as $O_1^{-1}(y,z) \stackrel{\text{def}}{=} \inf\{x : O_1(x,z) \geq y\}$.

Lemma 1 *Given EE improvement level, z, DR incentive rate, r, and DR event length, $h > 0$, the firm's optimal DR participation is to reduce power consumption by the proportion*

$$\alpha^{f*} = \min\{1, \frac{A(z,r)}{h}\}, \tag{8}$$

where $A(z,r) \stackrel{\text{def}}{=} O_1^{-1}(rP_z, z)$ is the maximum duration of a DR event (under given z and r) for which the firm will halt all production. The minimum cost per day, defined in (7), is

$$C(z,h,r) = cP_zT + \begin{cases} O(h,z) - rP_zh, & \text{if } h < A(z,r), \\ O(A(z,r),z) - rP_zA(z,r), & \text{if } h \geq A(z,r). \end{cases} \tag{9}$$

Note that shifting work to overtime results in marginal overtime cost, $O_1(\alpha h, z)$, and marginal DR benefit, rP_z. The firm would not curtail energy if the marginal overtime cost always exceeds the marginal incentive. Formally, if $O_1(0,z) \geq rP_z$, we have $A(z,r) = 0$ and thus $\alpha^{f*} = 0$.

When $O_1(0,z) < rP_z$, the optimal DR, α^{f*}, depends on the length of the DR event, h:

- If $h \leq A(z,r)$ (short DR event), the firm is willing to halt all production ($\alpha^{f*} = 1$) and reap the maximum DR benefit available.
- If $h > A(z,r)$ (long DR event), overtime costs are so high that it is optimal not to halt all production, but shut down a proportion that is equal to $A(z,r)/h$.

Next, we study the interaction between EE and DR. We first look into the effects of the EE decision on DR participation, as stated in the following proposition.

Proposition 2 *The effect of the EE improvement level, z, on the DR participation level, α^{f*}, depends on the overtime cost structure.*
(i) If EE improvement does not affect the overtime cost, i.e., $O(x,z) = O(x,0)\ \forall\ z \in [0,1)$, then the firm's optimal DR participation level, α^{f*}, (weakly) decreases in the EE improvement level, z.
(ii) If EE improvement reduces the overtime cost in such a way that $O(x,z) \equiv O(x(1-z),0)\ \forall\ z \in [0,1)$, then the firm's optimal DR participation level, α^{f*}, (weakly) increases in the EE improvement level, z.

In general, EE improvements affect the firm's DR participation level in two ways. First, EE improvements lower the baseline and, therefore, reduce the payment that the firm receives when curtailing for DR—this effect discourages DR participation. Second, EE investments may reduce the overtime cost, allowing work to be shifted at a lower cost, which encourages DR participation. The combined effect is that DR participation may increase or decrease in the EE improvement level, as seen in Proposition 2. In Proposition 2(i), only the first effect is present and, therefore, EE improvements reduce DR participation. In part (ii), both energy costs and the amount of overtime work are reduced by proportion z. Due to the convexity of the overtime cost function, marginal overtime costs decrease faster than the loss in marginal DR benefits, thereby increasing the DR participation level as EE improves.

5.1.2 Optimal EE Improvement

To decide the upfront EE investment, the firm aims to minimize the sum of the upfront investment cost and the discounted daily operating costs:

$$\min_{z\in[0,1)} \; I(z) + \mathbb{E}\left[\sum_{n=1}^{N} \delta_f^n C(z, H, R)\right] \;\equiv\; I(z) + \gamma^f\, \mathbb{E}\big[C(z, H, R)\big]. \tag{10}$$

With the knowledge of the minimal cost $C(z, h, r)$ from Lemma 1, we now solve for the optimal EE improvement level, denoted as z_{DR}^{f*}, for the firm. For all realistic parameter combinations, the objective in (10) is unimodal and the first-order condition uniquely determines the optimal EE improvement level. Furthermore, employing the supermodularity property, we can analyze the monotonicity property of z_{DR}^{f*}. The monotonicity property does not require uniqueness. In rare situations when the optimal solution is nonunique, the descending notion in Topkis (1978) can be used in lieu of the decreasing notion in Proposition 3. These results are summarized in the following proposition.

Proposition 3 *For the EE investment problem with DR in* (10)*, if the optimal EE improvement level z_{DR}^{f*} is greater than zero, then it satisfies*

$$I'(z_{\mathrm{DR}}^{f*}) + \gamma^f \mathbb{E}\left[\int_0^{A(z_{\mathrm{DR}}^{f*}, R)} \big(RP_0 + O_{12}(x, z_{\mathrm{DR}}^{f*})\big)\big(1 - F_H(x)\big)dx\right] - \gamma^f cP_0T = 0. \tag{11}$$

If $-O_{12}(x, z) \le O_1(x, z)/(1 - z)$, the EE improvement level, z_{DR}^{f*}, decreases when the DR incentive rate, R, is shifted or scaled upward. In particular, the EE improvement level in the presence of DR is lower than the EE improvement level in the absence of DR, i.e., $z_{\mathrm{DR}}^{f*} \le z^{f*}$.

As shown in (11), the optimal EE improvement, z_{DR}^{f*}, balances the marginal cost and the marginal benefit of EE. The marginal cost of EE includes not only the marginal upfront cost of installation, $I'(z_{\mathrm{DR}}^{f*})$, but also the marginal loss in DR revenue, $\gamma^f \mathbb{E}\Big[\int_0^{A(z_{\mathrm{DR}}^{f*}, R)} RP_0\big(1 - F_H(x)\big)dx\Big]$. The marginal benefit from more EE is a marginal reduction in electricity bills, $\gamma^f cP_0T$, as well as the reduced marginal overtime cost, $\gamma^f \mathbb{E}\Big[\int_0^{A(z_{\mathrm{DR}}^{f*}, R)} -O_{12}(x, z_{\mathrm{DR}}^{f*})\big(1 - F_H(x)\big)dx\Big]$. Comparing (11) to (2), we see that DR alters the original tradeoff by introducing the effect of EE investment on DR-related revenue and cost into the equation for determining the EE investment. Importantly, Proposition 3 reveals that the firm's opportunity to participate in DR discourages its EE investment, because EE installations inadvertently reduce the firm's DR revenue stream.

In Proposition 3, the sufficient condition is satisfied in most situations. To see this, consider a special overtime cost function $O(x,z) = O(x,0)(1-z)$, in which an EE improvement level of z also reduces overtime cost by proportion z. In this special case, it can be readily verified that $\frac{-O_{12}(x,z)}{O_1(x,z)} = \frac{1}{1-z}$, i.e., the sufficient condition holds with equality. In most realistic situations, as a side-benefit of EE improvements, the overtime cost would not decrease as fast as energy consumption, which amounts to the inequality $\frac{-O_{12}(x,z)}{O_1(x,z)} \le \frac{1}{1-z}$ stated in the proposition. Here, $\frac{1}{1-z}$ is the proportion of marginal energy savings ($\frac{-\partial P_z/\partial z}{P_z} = \frac{1}{1-z}$), and $\frac{-O_{12}(x,z)}{O_1(x,z)}$ is the proportion of marginal overtime cost savings due to EE improvement. We further remark that even if the above sufficiency condition does not hold, the firm's choice of EE improvement level, z, typically still decreases when the incentive rate increases for all but the most extreme cases of overtime cost functions. In §5.3, we will examine one overtime function that does not satisfy the sufficiency condition and for which the firm still decreases its EE installations as DR incentives increase.

5.2 Minimizing the Firm's Impact on Societal Costs

5.2.1 Optimal DR Under Given EE Improvement

We now optimize the DR participation level and EE improvement level at the firm from the societal perspective. During a DR event, the firm's curtailment helps reduce the peaking cost of electricity generation. Let M be the difference in the marginal generation costs between the peak and the off-peak:

$$M \equiv G_p - G_b. \tag{12}$$

Let m, g_p, and g_b denote the realization of M, G_p, and G_b in a day. One MWh of energy shifted from the peak to the off-peak saves M for society, which is balanced against the increase in overtime costs. The DR incentive payment and retail costs of electricity are transfer payments that are not included in the societal costs. The societal costs include the inconvenience costs to the workers who work overtime and have to delay or cancel personal activities. The overtime compensation (part of the overtime cost) is an approximation of these inconvenience costs, and thus not a transfer payment.

Similar to §5.1, we consider a sequential optimization problem. Knowing the EE improvement level, z, the DR event duration, h, and the two-level marginal generation costs, g_p and g_b, the firm's DR participation level, α, is chosen to minimize the sum of overtime cost and the marginal impact of the firm's energy use on society:

$$\begin{aligned} C^s(z,h,g_p,g_b) &= \min_{\alpha\in[0,1]} g_bP_z(T-h)+g_p(1-\alpha)P_zh+g_b\alpha P_zh+O(\alpha h,z) \\ &= g_bP_z(T-h)+g_pP_zh+\min_{\alpha\in[0,1]}\Big\{O(\alpha h,z)-m\alpha P_zh\Big\}. \end{aligned} \tag{13}$$

In (13), the first two terms represent the marginal generation costs of the firm's energy use without DR participation, $O(\alpha h,z)$ represents the overtime costs, and $-m\alpha P_zh$ represents the generation cost reduction due to DR participation.

The following lemma gives the optimal DR desired by society.

Lemma 2 *Given EE improvement level, z, marginal cost difference, m, and DR event length, $h>0$, the firm's DR participation level that is optimal for society is*

$$\alpha^{s*} = \min\{1, \frac{B(z,m)}{h}\}, \tag{14}$$

where $B(z,m) \overset{\text{def}}{=} O_1^{-1}(mP_z,z)$. The minimum impact per day, defined in (13), is

$$C^s(z,h,g_p,g_b) = g_bP_z(T-h)+g_pP_zh+\begin{cases} O(h,z)-mP_zh, & \text{if } h<B(z,m), \\ O(B(z,m),z)-mP_zB(z,m), & \text{if } h\geq B(z,m). \end{cases} \tag{15}$$

Similar to $A(z,r)$ in §5.1, $B(z,m)$ is the maximum duration of peak demand for which society desires the firm to halt all production. The structure of the optimal DR for society is parallel to that for the firm in Lemma 1. The only difference is that the marginal benefit of DR is mP_z to society, while it is rP_z to the firm.

Analogous to Proposition 2, the effect of EE improvement on the DR participation level depends on the overtime cost structure, as stated in the following proposition.

Proposition 4 *(i) If EE improvement does not affect the overtime cost, i.e., $O(x,z)=O(x,0)$, then the optimal DR participation level, α^{s*}, (weakly) decreases in the EE improvement level, z. (ii) If EE improvement reduces the overtime cost in such a way that $O(x,z)=O(x(1-z),0)$, then the optimal DR participation level, α^{s*}, (weakly) increases in the EE improvement level, z.*

Propositions 2 and 4 suggest that the dependence of DR participation on the EE improvement is qualitatively the same regardless of whether we are minimizing costs to the firm or to society.

5.2.2 Optimal EE Improvement

We next find the EE improvement at the firm desired by society. Parallel to the firm's problem in (10), the EE improvement desired by society is determined by solving

$$\min_{z\in[0,1)} \; I(z) + \mathbb{E}\left[\sum_{n=1}^{N} \delta_s^n C^s(z,H,G_p,G_b)\right] \;\equiv\; I(z) + \gamma^s\,\mathbb{E}\big[C^s(z,H,G_p,G_b)\big], \tag{16}$$

where $C^s(z,H,G_p,G_b)$ is given in (15) and the expectation is taken on H, G_p, and G_b.

Proposition 5 *For the EE investment problem with DR in* (16)*, if the optimal EE improvement level* $z_{\mathrm{DR}}^{s*} > 0$*, then it satisfies*

$$I'(z_{\mathrm{DR}}^{s*}) + \gamma^s\mathbb{E}\left[\int_0^{B(z_{\mathrm{DR}}^{s*},M)} \big(MP_0 + O_{12}(x,z)\big)\big(1 - F_H(x)\big)dx\right] = \gamma^s\bar{c}^s P_0 T, \tag{17}$$

where $\bar{c}^s$ *is the marginal cost of generating electricity averaged over* T*, as defined in* (5)*. If* $-O_{12}(x,z) \le O_1(x,z)/(1-z)$*, the EE improvement level in the presence of DR is lower than the EE improvement level in the absence of DR, i.e.,* $z_{\mathrm{DR}}^{s*} \le z^{s*}$*.*

Proposition 5 is the societal analog to Proposition 3. Importantly, Propositions 3 and 5 reveal that the presence of DR reduces the optimal EE investment level, regardless of whether we are minimizing costs to the firm or to society.

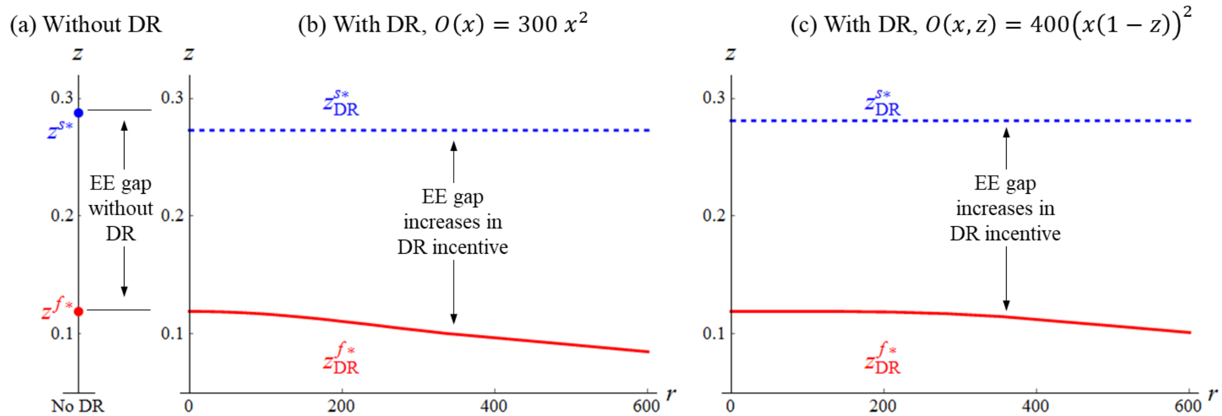
5.3 Energy Efficiency Gap In the Presence of Demand Response

In §4.3, the EE gap in the absence of DR is measured as $z^{s*} - z^{f*}$. With DR participation, the EE gap, measured as $z_{\mathrm{DR}}^{s*} - z_{\mathrm{DR}}^{f*}$, needs to be reassessed. The presence of DR reduces both the firm and societal choices of EE improvement level. The EE improvement level desired by society at the firm, z_{DR}^{s*} in Proposition 5, does not depend on the DR incentive rate. However, the EE improvement level desired by the firm, z_{DR}^{f*} in Proposition 3, decreases in the DR incentive rate. For ease of exposition, we let DR incentive R assume a deterministic value r and use the notation $z_{\mathrm{DR}}^{f*}(r)$ to emphasize the dependence of the firm's EE improvement level on r.

Proposition 6 *In the presence of DR, the EE gap,* $z_{\mathrm{DR}}^{s*} - z_{\mathrm{DR}}^{f*}(r)$*, increases in the DR incentive rate,* r*.*

Figure 2 depicts how DR affects the firm's choice and society's choice of EE improvement and the resulting EE gap, using an example with realistic parameters justified in the appendix. Figure 2(a) shows the existence of the EE gap in the absence of DR (Proposition 1). Figure 2(b) and (c) show that, despite the two very different overtime cost structures, the effect of DR incentive on the EE

Figure 2: Energy efficiency gap with and without the presence of DR



Firm's parameters: $P_0 = 10$ MW, $T = 16$ hours, $c = \$80/\text{MWh}$, $\delta_f = 0.9996$, $N = 1500$.
Societal parameters: $\delta_s = 0.9999$, $G_b = \$200/\text{MWh}$ (natural gas) and $G_p = \$600/\text{MWh}$ (petroleum).
EE investment cost: $I(z) = 50z^2/(1-z)$ million dollars.
DR event duration: $H = 0$ with probability 0.8, $H = 2, 3, 4, 5$ hours with equal probabilities 0.05.
See the appendix for detailed justifications of these parameters.

gap is qualitatively similar. When r increases, DR participation becomes more lucrative to the firm, thus reducing the incentive to invest in EE, widening the EE gap, regardless of the overtime cost structure. The scale of industrial firms makes their EE decision very impactful; in this example, if the firm were to operate 300 days/year, increasing z by 0.01 equates to a yearly energy savings of 480 MWh. This is roughly equivalent to the annual consumption of 45 U.S. households. Note that the overtime cost function in Figure 2(b) satisfies the sufficiency condition of Proposition 3, while the overtime function in Figure 2(c) does not.

For decades, policy makers have attempted to close the EE gap with incentives and regulations. Such efforts seem to be growing in scope, judging by the recent initiatives mentioned in §1. Our analysis suggests that policy makers should understand the impact of DR on the EE gap to design effective policies.

6. Coordinating Energy Efficiency and Demand Response

After Section 5 demonstrated the interactions between EE and DR, we are left with the question of how best to coordinate EE and DR to maximize their combined benefits to society. Intuitively, the firm's impact on societal costs would be minimized if the societal optimal levels of EE improvement and DR participation were induced at the firm. This would entail fully closing the EE gap while simultaneously aligning the DR incentive to the difference in peak and off-peak generation costs

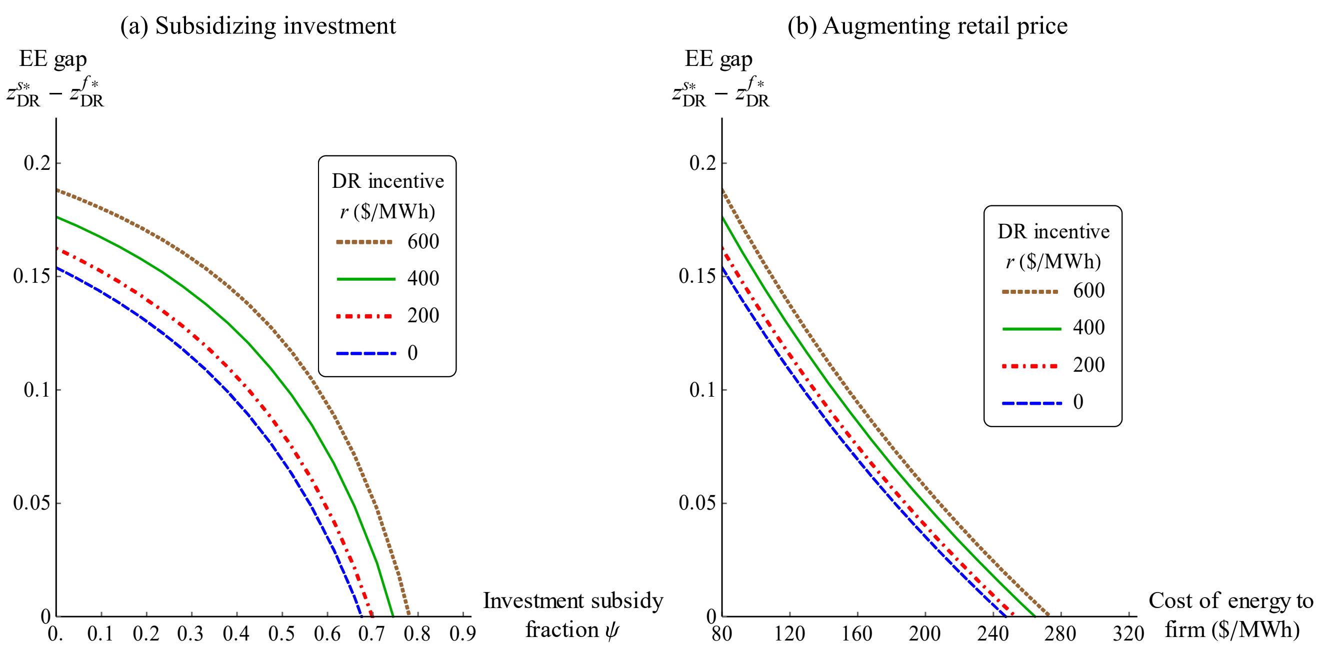
($R = M^+ = (G_p - G_b)^+$).

Proposition 7 *To minimize the firm's impact on societal costs, both EE improvement level and DR participation level must be aligned to the societal optimal levels (characterized in Lemma 2 and Proposition 5). The DR incentive should be the positive difference between the peak and off-peak cost of power generation:* $R = M^+ = (G_p - G_b)^+$.

How practical is it to align firm actions to the societal optimal? As we will see, full alignment is not practical for a variety of reasons. However, we can still endeavor to improve upon the status quo, in which policy makers design EE incentives without consideration of DR programs. In this section, we demonstrate the consequences of failing to consider the interactions between EE and DR and propose ways to coordinate the design of these two demand-side programs.

Figure 3: Closing the EE gap under different EE incentives in the presence of DR

Note: Same parameters as Figure 2 and $O(x, z) = 300x^2$.



Upon recognition of the EE gap, policy makers have tried to induce firms to increase their EE investments. A common incentive is to subsidize the cost of EE investments, so that the firm either pays less out of pocket or receives a rebate for investment costs incurred. For illustrative purposes, we consider a subsidy policy under which the government subsidizes proportion $\psi \in [0, 1]$ of the investment so that the firm only incurs upfront cost $(1 - \psi)I(z)$. Another frequently discussed policy is a carbon tax that augments the retail cost of electricity to help cover the environmental externalities of energy use. If the retail price of electricity is augmented to $\hat{c} > c$, the firm has more incentive to reduce energy use through EE improvements.

Fully closing the EE gap, as recommended in Proposition 7, is not typically a viable goal, as shown in Figure 3, which uses the same parameters as Figure 2(b). An EE subsidy in excess of 65% of investment cost or an electricity price in excess of $240/MWh is required to fully close the EE gap. It will frequently be the case that this high level of EE incentive is infeasible for various reasons, e.g., a high subsidy may not be practical in the government budget or the socially-desirable carbon tax is politically unpalatable. The results in this section are robust to the overtime cost structure. See Online Appendix A for details.

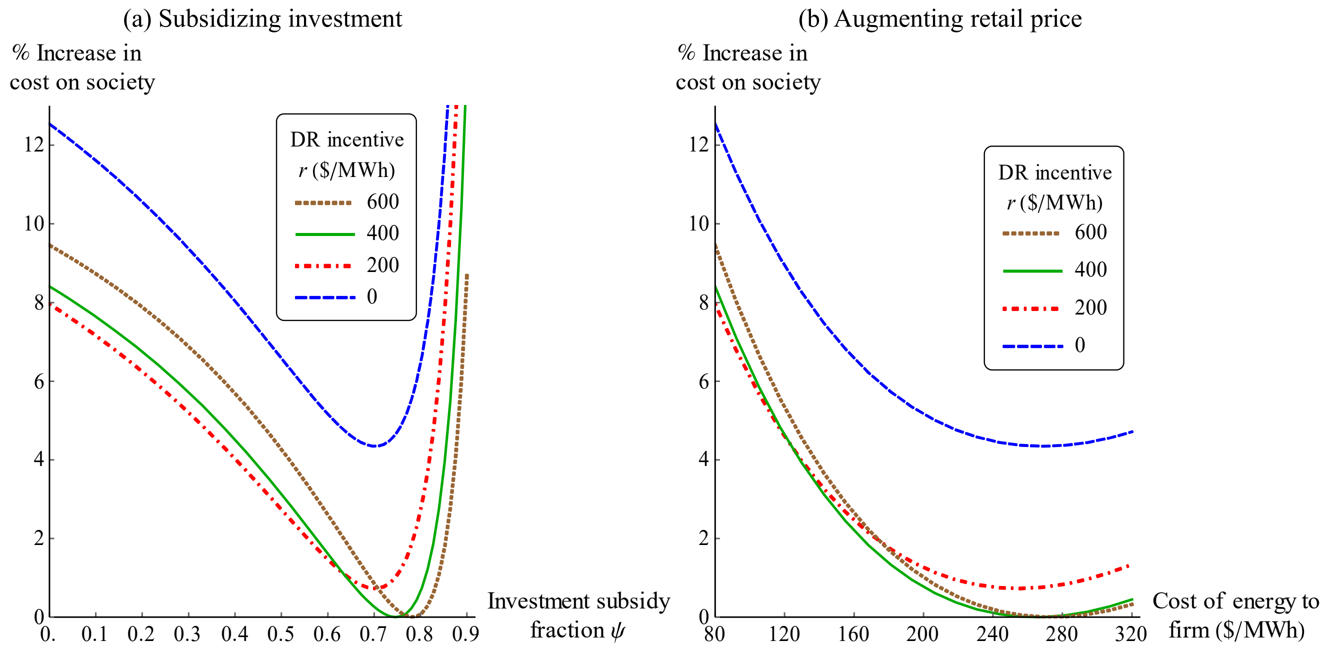
If it is impractical to fully close the gap, should our goal be to minimize the EE gap? For a given DR incentive, this is certainly a fine goal, as reducing the magnitude of the EE gap will then decrease the firm's impact on societal costs. However, the story is not as clear if we were to alter the DR incentive, which inevitably affects the EE gap. As the EE gap increases in the DR incentive (see Proposition 6), the EE gap is minimized when DR participation is eliminated by setting the DR incentive to zero. This is illustrated in Figures 3(a) and (b), where a given EE subsidy level or carbon tax level would reduce the EE gap the most if the DR incentive were set to $0. This is the level of EE gap that would be expected by creating EE incentives without consideration of DR. The presence of DR, however, renders this an overestimate of the firm's actual level of EE improvement. This is similar to $z^{f*} \geq z^{f*}_{\mathrm{DR}}$ in Proposition 3. We can test whether eliminating DR participation and the peak-savings that come with it would be problematic by studying the firm's impact on societal costs. Relative to the minimum possible costs of Proposition 7, Figure 4 shows the increase in cost on society due to the firm's electricity use for various combinations of EE and DR incentives.

In Figure 4, first note that a DR incentive of $0 always imposes a higher cost on society, regardless of the EE incentive, than a DR incentive of $200 or $G_p - G_b = \$400$. As such, we are not always justified in reducing the DR incentive to minimize the EE gap. Holding the EE incentive constant and reducing the EE gap by decreasing the DR incentive can increase costs to society.

As such, the EE gap is not the best metric of performance when firms also participate in DR. We must consider total societal costs, which are affected by both EE *and* DR incentives. In our example, when there is no DR incentive ($r = 0$) and no EE subsidy ($\psi = 0$) or taxes ($\hat{c} = c$), the firm's operations will impose 12.5% more cost on society than the minimum possible cost described in Proposition 7. If *only* DR incentives are introduced (refer to $\psi = 0$ in Figure 4(a) and $c = 80$ in Figure 4(b)), the firm's impact can be reduced to about 8% above the minimum possible cost. If *only* EE incentives are introduced (refer to $r = 0$ in Figure 4), the firm's impact can be reduced

Figure 4: Increased cost to society under different EE and DR incentives

Note: Same parameters as Figure 2 and $O(x, z) = 300x^2$.



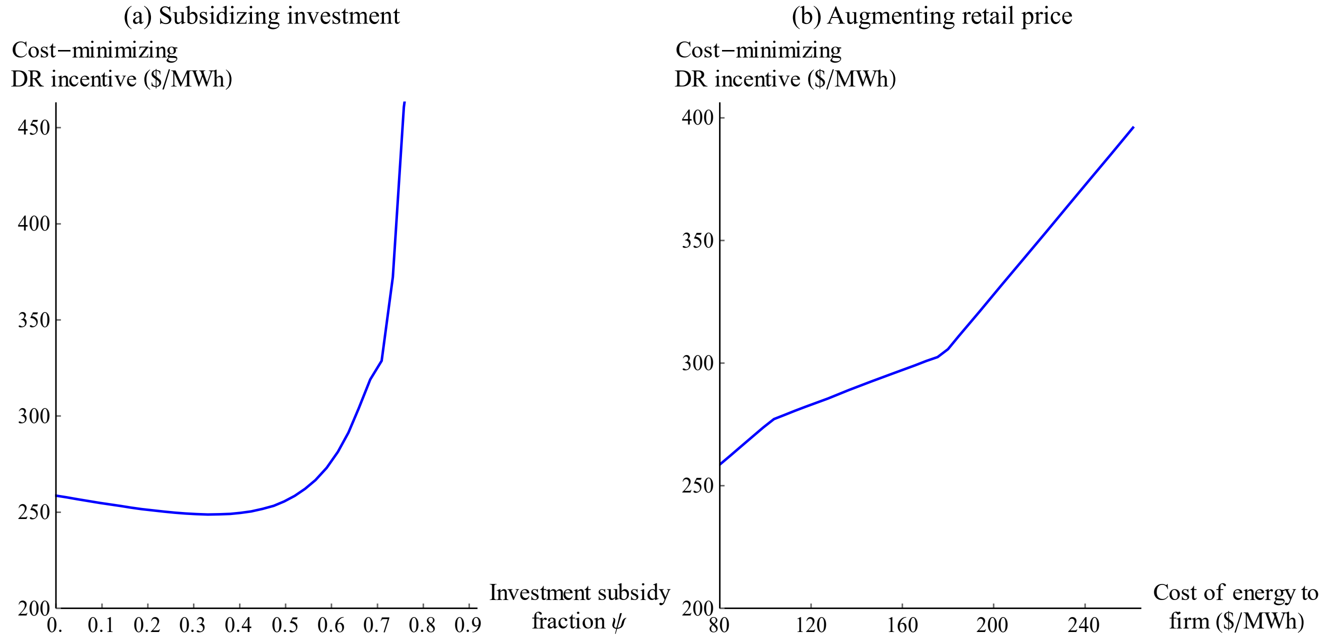
to about 4.5% above the minimum possible cost. Only coordinated EE *and* DR incentives can minimize the costs imposed on society. In this example, when 12.5% of the cost on society is saved by using the optimal EE and DR program offerings, the EE program contributes about two-thirds of the cost reduction, while the DR program contributes about one-third.

In addition to high levels of EE incentives being impractical, high levels of DR incentives may not be reasonable. Our societal model considers the environmental externalities of electricity generation. Depending on environmental regulations, utilities and third-party demand response providers, which administer DR programs, may not experience such costs. These DR program administrators may offer a DR incentive that is no larger than the difference in market clearing wholesale electricity price between peak and off-peak periods.

In cases of constrained incentive values, our model provides a framework to perform a constrained optimization. Specifically, we can find the best DR incentive for a given, feasible level of EE incentive (or vice versa). Figure 5 shows the case of finding the DR incentive that minimizes societal costs for a given EE incentive. Figure 5 reveals that, although $r = G_p - G_b = \$400$/MWh is part of the optimal incentive scheme, $r = \$400$/MWh is not necessarily optimal for every EE incentive level. In Figure 5(a), if the subsidy fraction is constrained by $\psi \leq 0.5$, then the best DR incentive is about \$250/MWh, instead of \$400/MWh. Figure 5(b) repeats this analysis for a

Figure 5: Cost-minimizing DR incentive under different EE policies

Note: Same parameters as Figure 2 and $O(x, z) = 300x^2$.



carbon tax policy that augments the cost of retail electricity. Note that the best DR incentive rate is correlated with the augmented cost of energy. The above findings strongly suggest that focusing on reducing the EE gap alone is not sufficient. Failure to coordinate EE and DR incentives can lead to significant increases in societal costs.

Finally, we comment on the case of $G_b > G_p$, where G_b and G_p may represent the marginal costs, including externalities, of coal-fired and natural gas-fired generators, respectively. Our model is general enough to include this case: Proposition 7 prescribes that the DR incentive should be zero, implying that the firm would not shift its production from peak to off-peak. And only in this case, closing the EE gap minimizes the firm's impact on society. With many coal-fired power plants retiring or being repowered with natural gas, $G_b < G_p$ stands to become even more prevalent in the future.

7. Conclusion and Extensions

In this paper, we have studied the interactions between long-term energy efficiency (EE) upgrades and daily demand response (DR) participation. We have demonstrated that the presence of a DR incentive will decrease the upfront EE investment. Both EE and DR play a role in modifying load on the electrical grid, and by understanding the desirable EE investment and DR participation from both the firm's and societal perspectives, we may determine more effective policies for demand-side management. While revisiting the EE gap caused by the environmental externalities of energy generation, we note that the gap must be reassessed in the presence of DR, because the opportunity to participate in DR affects the EE level desired by the firm. The EE gap increases with the magnitude of the DR incentive.

In the presence of DR, closing the EE gap at the firm may not minimize its impact on society. Thus, policy makers should not focus on reducing the EE gap alone, but aim to jointly design EE and DR programs to deliver maximum benefit. The framework in this paper provides guidance toward designing incentives to coordinate EE investments and DR participation. Most likely, these insights have not been considered by the policy makers. Without understanding these relations, policies designed to increase EE investment, such as investment subsidies and carbon taxes, may fail to achieve their desired outcome.

To make our model tractable and obtain insights into the basic tradeoffs between EE and DR, our analysis uses simplifying assumptions on the electricity generation costs on society. Although our bi-level cost structure does not capture the full spectrum of reality, it approximates the fact that, while cheaper power generators are sufficient to meet most demand, there are times when expensive peak generators are engaged to match demand peaks.

There may be other ways that a firm could gain more flexibility to respond to DR events, such as revising production targets. With more flexibility, we expect the firm to participate more in DR and thus gain a higher DR revenue. Consequently, the firm will have less incentive to invest in EE, which strengthens the results in the paper.

This paper does not model all of the potential benefits of EE and DR. For the electricity grid, demand-side management can defer or eliminate the need for infrastructure investment (Vine 2008). For a manufacturing firm, EE improvement may shield the firm against regulatory risk. Furthermore, implementation of EE and DR can demonstrate the firm's commitment to the environment—a growing requirement from large retailers (Walmart 2019).

In closing, we consider two extensions to our analysis. The first focuses on using an on-site generator to lower the firm's demand from the grid, instead of shifting production to overtime. The second considers incentive-based DR programs with mandatory curtailment. Our modeling framework can be readily generalized to these situations, and our main insights continue to hold.

On-Site Generation: Some firms perform DR not by turning off machines and working overtime, but instead by running an on-site generator. The on-site option provides some or all of the electricity needed by the firm, and the firm's demand on the external grid is thereby lessened. This option could be incorporated by altering the "overtime" costs in our model. Instead of $O(\cdot)$ being convex increasing, it would be linear, representing the linearly increasing cost of fueling the on-site generator as more load is switched to the on-site option. We can show that the on-site option will be run in an all-or-nothing way during DR events. Either it is economical for the firm to use the on-site generator to take on as much of the firm's load as possible, or it is not worthwhile to use the on-site option at all, based on the DR incentive. Policy makers are wary of using on-site generation to perform DR. A small-scale, on-site generator may be even more polluting than peaking power plants. As such, regulations often dictate what types of on-site generators may be used during DR events.

DR Programs with Mandatory Curtailment: These programs offer predetermined payments to a firm for committing to reduce consumption by a pre-specified amount in all future DR events. The firm receives this payment whether or not any events are called and receives an extra incentive payment, akin to R, per MWh curtailed when an event is called. The firm is penalized for not curtailing to their contracted level. The problem setup and notations are similar to the model in §3-§5. The main difference is that the firm (or society) now simultaneously optimizes the EE improvement level, z, and the DR participation level, α. In Online Appendix C, we show that the insights in the paper continue to hold. Notably, the firm's upfront EE investment decreases in the DR incentive, and the size of the EE gap still depends on the DR incentive.

Appendix: Numerical Settings

The base case considers a typical energy-intensive manufacturing firm that consumes $P_0 = 10$ MW during $T = 16$ operating hours. The firm's electricity cost is fixed at $c = \$80$ per MWh or 8 cents per kWh. The firm's daily discount factor is $\delta_f = 0.9996$, which is equivalent to 11.3% annual rate of return, assuming 300 working days per year. The planning horizon is 5 years or $N = 1500$ days.

When minimizing the firm's impact on societal costs, we use a daily discount factor $\delta_s = 0.9999$, which is equivalent to 3.0% annual discount rate. To estimate the environmental externalities of power generation, we refer to the emission factors for three typical power generating units in Table 1.

Monetizing environmental externalities is challenging and prior research tends to monetize a portion of the damages and give a wide range of estimates. The National Research Council (2010) estimates that the damage associated with three air pollutants—SO_2, NOx, and particulates—from coal-fired generation varies from 0.5 to 13 cents per kWh (5th and 95th percentiles, respectively). For natural gas-fired power plants, this damage is only 0.001 to 0.55 cents per kWh (5th and 95th percentiles, respectively). The National Research Council (2010) also provides estimates of amounts of green-house gas emissions and other pollutants including metals, radionuclides, effluents,

Table 1: Pounds of air pollutants produced per MWh of electricity generated

	Natural gas	Oil	Coal
Carbon dioxide	878	1804	2184
Carbon monoxide	0.30	0.36	2.18
Nitrogen oxides	0.69	4.93	4.80
Sulfur dioxide	0.005	12.34	27.21
Particulates	0.053	0.92	28.81
Mercury	0	7.7e-5	1.7e-4

Note: Compiled using data from EIA (1999, Table 2) and assuming the heat rate for natural gas, oil, and coal-fired units are 7,500, 11,000, 10,500 Btu per kWh, respectively. This table is for illustration of emissions from typical generation units. The vintage of the units and the equipped emission controls greatly influence the emission intensities.

and solid wastes occurring during resource extraction, transportation, and power generation, but comments that it is difficult to monetize the damages of these pollutants.

In the base case in this paper, the marginal cost of electricity generation including environmental externalities is $G_b = \$200$ per MWh for natural gas-fired generators and $G_p = \$600$ per MWh for oil-fired generators. We have conducted robustness tests for various combinations of parameters, and found that qualitative insights remain the same.

To reduce the firm's power consumption by a fraction z, we assume the required investment cost is $I(z) = \dfrac{50z^2}{1-z}$ million dollars. Nguyen et al. (2018) assume that energy saving is increasing and concave in the investment level I in a square root form, i.e., I is quadratic in energy savings. Our investment function $I(z)$ is approximately quadratic at low levels of EE improvement, and we introduce a denominator $1 - z$ to capture the reality that EE measures cannot eliminate energy use entirely. Our model results in a required investment of approximately \$5,000 for the first 1% improvement in EE, \$132,000 for 5% EE improvement, \$2.5 million for 20%, and so on.

In our model, we assume that the DR event duration has the following discrete distribution: $H = 0$ with probability 0.8, $H = 2, 3, 4, 5$ hours with probabilities 0.05. In other words, the expensive peaking generators will be used on average $\bar{H} = 0.7$ hours a day, or 3% of the time, which is a realistic capacity factor for peaking power plants.

Different DR programs involve different values of r, which typically varies between \$100 to \$500 per MWh. For example, the average payout for the Economic DR Program of the PJM Interconnection was \$121/MWh in 2014 (McAnany 2015). In our numerical setting, we allow the DR incentive to vary between 0 and \$600/MWh.

The firm responds to DR by delaying production. Our numerical analysis tests two forms of overtime cost structures:

(a) $O(x, z) = 300x^2$, in which EE improvement has no impact on overtime cost, and

(b) $O(x, z) = 400(x(1 - z))^2$, in which the EE improvement reduces the amount of overtime work and energy consumption by the same proportion z.

These two forms represent extreme cases: either EE saves nothing in labor costs or it saves as high of a percentage of labor costs as it does in energy costs. The numerical analysis in §6 focuses on the form in (a) above. The results for the overtime cost in (b) are shown in Online Appendix A.

References

Aflaki, S., P. R. Kleindorfer, V. S. de Miera Polvorinos 2013. Finding and implementing energy efficiency projects in industrial facilities. *Production and Operations Management* **22**(3) 503–517.

Albadi, M. H., E. El-Saadany 2008. A summary of demand response in electricity markets. *Electric Power Systems Research* **78**(11) 1989–1996.

Ata, B., A. S. Duran, O. Islegen 2018. An analysis of time-based pricing in retail electricity markets. *Available at SSRN 2826055*.

Boyd, G. A., J. X. Pang 2000. Estimating the linkage between energy efficiency and productivity. *Energy policy* **28**(5) 289–296.

California ISO 2013. Demand response and energy efficiency roadmap: Maximizing preferred resources. Available at https://www.caiso.com/Documents/DR-EERoadmap.pdf.

Chao, X., F. Y. Chen 2005. An optimal production and shutdown strategy when a supplier offers an incentive program. *Manufacturing & Service Operations Management* **7**(2) 130–143.

Daniels, K. M., R. Lobel 2014. Demand response in electricity markets: Voluntary and automated curtailment contracts. *Available at SSRN 2505203*.

DeCanio, S. J. 1993. Barriers within firms to energy-efficient investments. *Energy Policy* **21**(9) 906–914.

Duke Energy 2017. PowerShare. www.duke-energy.com/business/products/powershare. Accessed: 2019-05-30.

Energy Information Administration 1999. Natural gas 1998: Issues and trends. DOE/EIA-0560(98), www.eia.gov/oil_gas/natural_gas/analysis_publications/natural_gas_1998_issues_and_trends/it98.html.

Energy Information Administration 2012. Cheaper natural gas alters generation dispatch in Southeast. www.eia.gov/todayinenergy/detail.php?id=9090. Accessed: 2017-08-05.

Energy Information Administration 2016. Electric power sales, revenue, and energy efficiency Form EIA-861 detailed data files. www.eia.gov/electricity/data/eia861/. Accessed: 2017-08-12.

Energy Information Administration 2017. Form EIA-861: Annual Electric Power Industry Report – Instructions. www.eia.gov/survey/form/eia_861/instructions.pdf. Accessed: 2017-07-12.

Epstein, P. R., J. J. Buonocore, K. Eckerle, M. Hendryx, B. M. Stout III, R. Heinberg, R. W. Clapp, B. May,

N. L. Reinhart, M. M. Ahern, S. K. Doshi, L. Glustrom 2011. Full cost accounting for the life cycle of coal. *Annals of the New York Academy of Sciences* **1219**(1) 73–98.

Fernandez, M., L. Li, Z. Sun 2013. "Just-for-Peak" buffer inventory for peak electricity demand reduction of manufacturing systems. *International Journal of Production Economics* **146**(1) 178–184.

Gerarden, T. D., R. G. Newell, R. N. Stavins 2017. Assessing the energy-efficiency gap. *Journal of Economic Literature* **55**(4) 1486–1525.

Gillingham, K., R. G. Newell, K. Palmer 2009. Energy efficiency economics and policy. *Annual Review of Resource Economics* **1** 597–620.

Goldman, C., M. Reid, R. Levy, A. Silverstein 2010. Coordination of energy efficiency and demand response. Available from National Action Plan for Energy Efficiency, `www.epa.gov/energy/national-action-plan-energy-efficiency`.

Hirst, E., M. Brown 1990. Closing the efficiency gap: barriers to the efficient use of energy. *Resources, Conservation and Recycling* **3**(4) 267–281.

Hledik, R., J. Chang, R. Lueken 2016. The hidden battery: Opportunities in electric water heating. *The Brattle Group, Inc.*

Jaffe, A. B., R. N. Stavins 1994. The energy-efficiency gap: What does it mean? *Energy Policy* **22**(10) 804–810.

Jewell, W. 2014. The effects of residential energy efficiency on electric demand response programs. *System Sciences (HICSS), 2014 47th Hawaii International Conference on*. IEEE. pp. 2363–2372.

King, C., D. Delurey 2005. Efficiency and demand response: Twins, siblings, or cousins? *Public Utilities Fortnightly* **143**(3) 54–61.

Kök, A. G., K. Shang, Ş. Yücel 2016. Impact of electricity pricing policies on renewable energy investments and carbon emissions. *Management Science* **64**(1) 131–148.

Lohmann, T., S. Rebennack 2017. Tailored benders decomposition for a long-term power expansion model with short-term demand response. *Management Science* **63**(6) 2027–2048.

McAnany, J. 2015. 2014 demand response operations markets activity report. Available at `https://www.pjm.com/~/media/markets-ops/dsr/2014-demand-response-activity-report.ashx`. Accessed 2019-06-02.

Mohagheghi, S., N. Raji 2015. Dynamic demand response: A solution for improved energy efficiency for industrial customers. *Industry Applications Magazine, IEEE* **21**(2) 54–62.

Mooney, C. 2015. Obama just released the biggest energy efficiency rule in U.S. history. Washington Post, December 17, 2015.

Muthulingam, S., C. J. Corbett, S. Benartzi, B. Oppenheim 2013. Energy efficiency in small and medium-sized manufacturing firms: order effects and the adoption of process improvement recommendations. *Manufacturing & Service Operations Management* **15**(4) 596–615.

Nadel, S. 2017. Demand response programs can reduce utilities' peak demand an average of 10%, complementing savings from energy efficiency programs. *American Council for an Energy-Efficient Economy.*

Available from http://aceee.org/blog/2017/02/demand-response-programs-can-reduce.

National Research Council 2010. *Hidden costs of energy: unpriced consequences of energy production and use*. National Academies Press.

Natural Resources Canada 2006. Improving energy performance in Canada. Available at oee.nrcan.gc.ca/publications/statistics/parliament05-06/chapter5.cfm.

New York Independent System Operator 2016. NYISO demand response programs: Frequently asked questions for prospective resources. www.nyiso.com/public/markets_operations/market_data/demand_response/index.jsp. Accessed: 2017-08-22.

NextEra Energy Services 2017. Real-time price response. www.nexteraenergyservices.com/for-business/business-solutions/demand-response/demand-response-products. Access: 2019-05-30.

Nguyen, J., K. Donohue, M. Mehrotra 2018. Closing a supplier's energy efficiency gap through assessment assistance and procurement commitment. *Management Science* **65**(1) 122–138.

O'Connell, N., P. Pinson, H. Madsen, M. O'Malley 2014. Benefits and challenges of electrical demand response: A critical review. *Renewable and Sustainable Energy Reviews* **39** 686–699.

Olsen, D., S. Goli, D. Faulkner, A. McKane 2010. Opportunities for energy efficiency and demand response in the california cement industry. *Ernest Orlando Lawrence Berkeley National Laboratory Report LBLN-4849E*.

PJM Interconnection 2017. Economic demand resource in the energy market. www.pjm.com/Globals/Training/Courses/v-eco-dsr.aspx. Accessed: 2019-05-30.

Sandberg, P., M. Söderström 2003. Industrial energy efficiency: the need for investment decision support from a manager perspective. *Energy Policy* **31**(15) 1623–1634.

Scott, D., R. Castillo, K. Larson, B. Dobbs, D. Olsen 2015. Refrigerated warehouse demand response strategy guide. *Ernest Orlando Lawrence Berkeley National Laboratory Report LBNL-1004300*.

Siano, P. 2014. Demand response and smart grids—a survey. *Renewable and Sustainable Energy Reviews* **30** 461–478.

Smith, K., K. Managan 2012. Energy efficiency and demand response: Working together in an integrated approach to managing energy. Institute for Building Efficiency, Johnson Controls.

Strbac, G. 2008. Demand side management: Benefits and challenges. *Energy Policy* **36**(12) 4419–4426.

Topkis, D. M. 1978. Minimizing a submodular function on a lattice. *Operations Research* **26**(2) 305–321.

U.S. Department of Energy 2006. Benefits of demand response in electricity markets and recommendations for achieving them. *U.S. Department of Energy, Washington, DC, USA, Tech. Rep.*

Vine, E. 2008. Breaking down the silos: the integration of energy efficiency, renewable energy, demand response and climate change. *Energy Efficiency* **1**(1) 49–63.

Walmart 2019. 2019 Walmart environmental, social, and governance report. Available from https://corporate.walmart.com/esgreport/. Accessed: 2019-07-30.

White House Council of Economic Advisors 2016. A retrospective assessment of clean energy investments in the Recovery Act. Available at `www.whitehouse.gov/administration/eop/cea`.

Worrell, E., J. A. Laitner, M. Ruth, H. Finman 2003. Productivity benefits of industrial energy efficiency measures. *Energy* **28**(11) 1081–1098.

York, D., M. Kushler 2005. Exploring the relationship between demand response and energy efficiency: A review of experience and discussion of key issues. *Technical report*. American Council for an Energy-Efficient Economy.

Online Appendices

A. Results Under Alternative Overtime Cost Structure

This section shows the results under the overtime cost $O(x, z) = 400(x(1-z))^2$, which implies that the EE improvement reduces the amount of overtime work and energy consumption by the same proportion z. The EE gap under this cost structure is shown in Figure 2(c) in the paper. Here, we present results parallel to Figures 3 and 4 in the paper.

Figure A.1: Closing the EE gap under different EE incentives in the presence of DR $O(x, z) = 400(x(1-z))^2$.

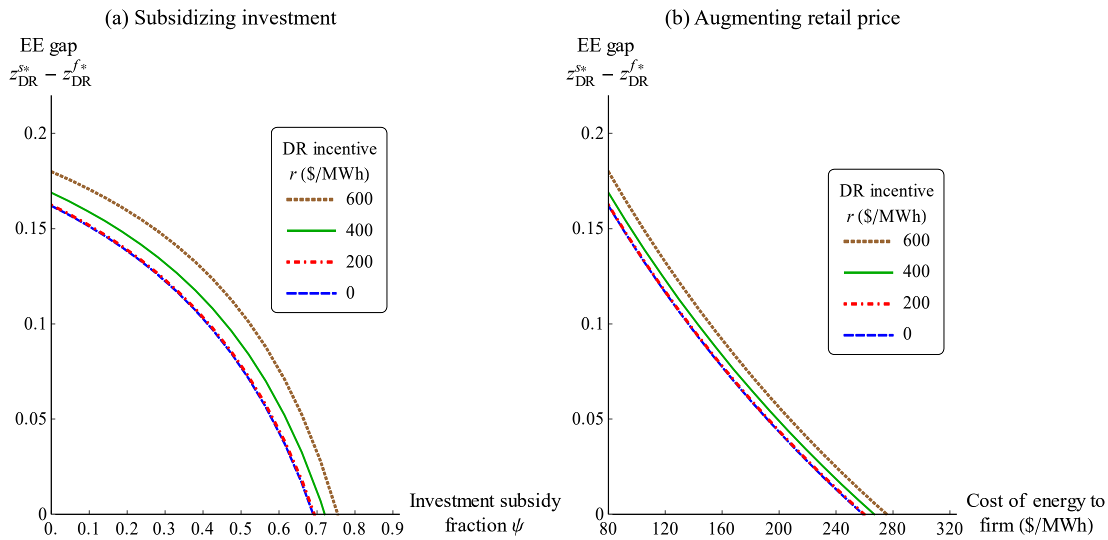
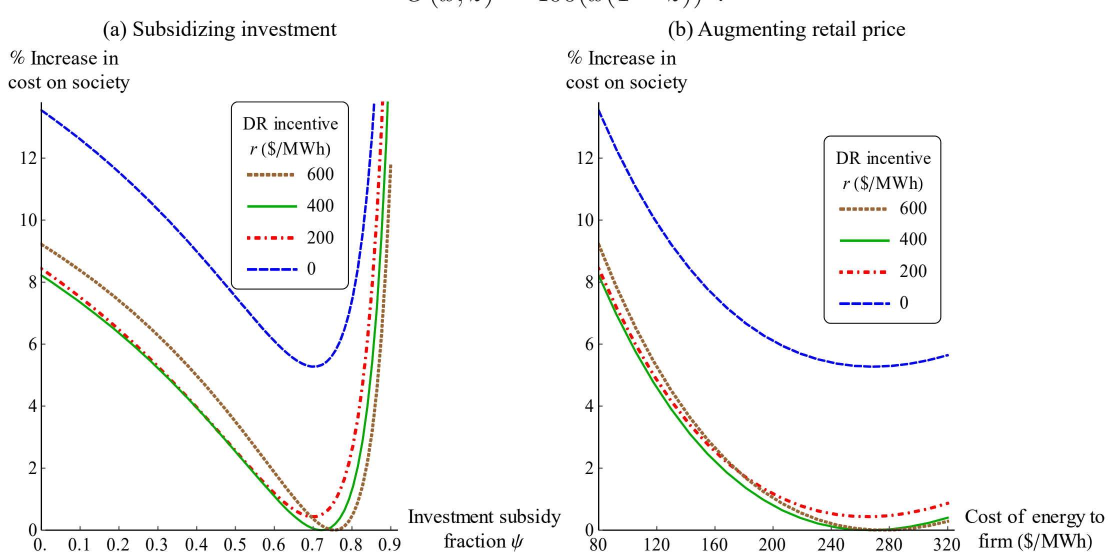


Figure A.2: Increased cost to society under different EE and DR incentives $O(x, z) = 400(x(1-z))^2$.



From Figures A.1 and A.2, we see that the qualitative insights are robust with respect to the overtime cost structures. In particular, Figure A.1 confirms that a higher DR incentive leads to a wider EE gap. Figure A.2 supports the finding that both EE and DR programs contribute significantly to the reduction of the cost on society.

B. Proofs

Proof of Proposition 1: Comparing (2) and (6), we see that both the firm and society face the same convex investment cost function, $I(\cdot)$. Thus, $z^{s*} > z^{f*}$ if $\gamma^s \bar{c}^s > \gamma^f c$. ∎

Proof of Lemma 1: We solve the minimization problem in (7): $\min_{\alpha \in [0,1]} \{O(\alpha h, z) - r\alpha P_z h\}$. Because the objective function is convex in α, the optimal solution can be determined by the first-order condition and expressed as

$$\alpha^{f*} = \begin{cases} 0, & \text{if } O_1(0,z) \geq rP_z, \\ O_1^{-1}(rP_z, z)/h = A(z,r)/h, & \text{if } O_1(0,z) < rP_z < O_1(h,z), \\ 1, & \text{if } O_1(h,z) \leq rP_z. \end{cases} \tag{A.1}$$

In the first case in (A.1), because $O_1(0,z) \geq rP_z$, we have $O_1^{-1}(rP_z, z) = 0$ by the definition of $O_1^{-1}(y,z)$ and, thus, $A(z,r) = 0$. In the last case, $O_1(h,z) \leq rP_z$ implies that $A(z,r)/h = O_1^{-1}(rP_z, z)/h \geq 1$. Combining the three cases, we have $\alpha^{f*} = \min\{1, A(z,r)/h\}$, which is (8).

Substituting the optimal DR into (7), we can verify the minimum cost expressed in (9). ∎

Proof of Proposition 2: ***(i) The case of*** $O(x,z) = O(x,0)$. Because $O_1(x,0)$ is continuous and strictly increasing in x (assumed in §3.2), $O_1^{-1}(y,0)$ increases in y. Therefore, $A(z,r) = O_1^{-1}(rP_z, 0) = {O_1}^{-1}(r(1-z)P_0, 0)$ decreases in z and hence α^{f*} decreases in z.

(ii) The case of $O(x,z) = O(x(1-z),0)$. In this case, we have $O_1(x,z) = (1-z)O_1(x(1-z),0)$ and $O_1^{-1}(y,z) = \inf\{x : O_1(x(1-z),0) \geq y/(1-z)\} = \frac{O_1^{-1}(y/(1-z),0)}{1-z}$. Then, $A(z,r) = O_1^{-1}(rP_z, z) = \frac{O_1^{-1}(rP_0, 0)}{1-z}$. When z increases, $A(z,r)$ increases and, therefore, α^{f*} increases in z. ∎

Proof of Proposition 3: ***(i) First-order condition.*** Because $I(z) \to \infty$ as $z \to 1$, we must have $z_{\text{DR}}^{f*} < 1$. Thus, if $z_{\text{DR}}^{f*} > 0$, it must be an interior solution and satisfies the first-order condition. It suffices to prove that the first-order derivative of the objective in (10) with respect to z is

$$I'(z) - \gamma^f cP_0 T + \gamma^f \int_{r_{\min}}^{r_{\max}} \left[\int_0^{A(z,r)} \big(rP_0 + O_{12}(x,z)\big)\big(1 - F_H(x)\big)dx \right] dF_R(r), \tag{A.2}$$

for all $z \in (0,1)$.

Using (9), the problem in (10) can be written as

$$\min_{z\in[0,1)} C^f(z) \equiv I(z) + \gamma^f cP_z T + \gamma^f \int_{r_{\min}}^{r_{\max}} L(z,r)dF_R(r), \tag{A.3}$$

where

$$L(z,r) = \int_0^{A(z,r)} \big(O(x,z) - rP_z x\big)dF_H(x) + \big(O(A(z,r),z) - rP_z A(z,r)\big)\big(1 - F_H(A(z,r))\big). \tag{A.4}$$

Consider three cases. First, when $O_1(0,z) < rP_z < O_1(h_{\max}, z)$, we have

$$\begin{aligned}
\frac{\partial L}{\partial z} &= \int_0^{A(z,r)} \big(O_2(x,z) + rP_0 x\big)dF_H(x) \\
&\quad + \Big(O_1(A(z,r),z)A'(z) + O_2(A(z,r),z) + rP_0 A(z,r) - rP_z A'(z)\Big)\big(1 - F_H(A(z,r))\big) \\
&= rP_0\Big[\int_0^{A(z,r)} x dF_H(x) + A(z,r)\big(1 - F_H(A(z,r))\big)\Big] \\
&\quad + \Big[\int_0^{A(z,r)} O_2(x,z)dF_H(x) + O_2(A(z,r),z)\big(1 - F_H(A(z,r))\big)\Big] \\
&= rP_0 \int_0^{A(z,r)} \big(1 - F_H(x)\big)dx + \int_0^{A(z,r)} \big(1 - F_H(x)\big)O_{12}(x,z)dx, \\
&= \int_0^{A(z,r)} \big(rP_0 + O_{12}(x,z)\big)\big(1 - F_H(x)\big)dx,
\end{aligned} \tag{A.5}$$

where the first equality applies the Leibniz integral rule (with some terms canceled), the second equality follows from $O_1(A(z,r),z) = rP_z$ because $A(z,r) = O_1^{-1}(rP_z, z)$ by definition, and the third equality employs integration by parts. Using (A.5), the derivative of (10) with respect to z in this case is exactly (A.2).

Second, when $rP_z \geq O_1(h_{\max}, z)$, we have $A(z,r) = O_1^{-1}(rP_z, z) \geq h_{\max}$. Because $F_H(x) = 1$ for $x \geq h_{\max}$, we have

$$L(z,r) = \int_0^{h_{\max}} \big(O(x,z) - rP_z x\big)dF_H(x), \tag{A.6}$$

and the first-order derivative is

$$\begin{aligned}
\frac{\partial L}{\partial z} &= \int_0^{h_{\max}} \big(O_2(x,z) + rP_0 x\big)dF_H(x) \\
&= \big(O_2(h_{\max}, z) + rP_0 h_{\max}\big) - \int_0^{h_{\max}} \big(O_{12}(x,z) + rP_0\big)F_H(x)dx \\
&= \int_0^{h_{\max}} \big(O_{12}(x,z) + rP_0\big)(1 - F_H(x))dx.
\end{aligned}$$

We can replace $h_{\max}$ in the last equation by $A(z,r) \geq h_{\max}$ because $F_H(x) = 1$ for $x \geq h_{\max}$. Hence, (A.2) again holds.

Third, when $rP_z \le O_1(0,z)$, we have $A(z,r)=0$ and the firm will not participate in DR, and the objective in (A.3) reduces to $C^f(z) = I(z) + \gamma^f cP_zT$, which coincides with the objective in (1) when EE is considered in isolation. The first order derivative is exactly (A.2), noting that $A(z,r)=0$ in this case.

(ii) Monotonicity. Consider first scaling incentive rate R by a scaling factor $s>0$. That is, the random variable sR has a distribution function $F_{sR}(r) = F_R(r/s)$. Denote the firm's total cost function in (10) as $C^f(z,s)$. The derivative in (A.2) becomes

$$\begin{aligned}\frac{\partial C^f}{\partial z} &= I'(z) - \gamma^f cP_0T + \gamma^f \int_{r_{\min}}^{r_{\max}} \left[\int_0^{A(z,r)} \big(rP_0 + O_{12}(x,z)\big)\big(1-F_H(x)\big)dx\right] dF_{sR}(r),\\ &= I'(z) - \gamma^f cP_0T + \gamma^f \int_{r_{\min}}^{r_{\max}} \left[\int_0^{A(z,sr)} \big(srP_0 + O_{12}(x,z)\big)\big(1-F_H(x)\big)dx\right] dF_R(r). \qquad \text{(A.7)}\end{aligned}$$

The derivative of (A.7) with respect to s is

$$\begin{aligned}\frac{\partial^2 C^f}{\partial z \partial s} &= \gamma^f \int_{r_{\min}}^{r_{\max}} \left[\big(srP_0 + O_{12}(A(z,sr),z)\big)\big(1-F_H(A(z,sr))\big)A_2(z,sr)r + \int_0^{A(z,sr)} rP_0\big(1-F_H(x)\big)dx\right] dF_R(r)\\ &= \gamma^f \int_{r_{\min}}^{r_{\max}} \left[\big(rP_0 + O_{12}(A(z,r),z)\big)\big(1-F_H(A(z,r))\big)A_2(z,r) + \int_0^{A(z,r)} P_0\big(1-F_H(x)\big)dx\right] r/s \cdot f_R(r/s)dr\end{aligned}$$
(A.8)

Next, consider shifting R by a distance of $s>0$. That is, $R+s$ has a distribution function $F_{R+s}(r) = F_R(r-s)$. The derivative in (A.2) becomes

$$\begin{aligned}\frac{\partial C^f}{\partial z} &= I'(z) - \gamma^f cP_0T + \gamma^f \int_{r_{\min}}^{r_{\max}} \left[\int_0^{A(z,r)} \big(rP_0 + O_{12}(x,z)\big)\big(1-F_H(x)\big)dx\right] dF_{R+s}(r),\\ &= I'(z) - \gamma^f cP_0T + \gamma^f \int_{r_{\min}}^{r_{\max}} \left[\int_0^{A(z,r+s)} \big((r+s)P_0 + O_{12}(x,z)\big)\big(1-F_H(x)\big)dx\right] dF_R(r).\end{aligned}$$
(A.9)

The derivative of (A.9) with respect to s is

$$\begin{aligned}\frac{\partial^2 C^f}{\partial z \partial s} &= \gamma^f \int_{r_{\min}}^{r_{\max}} \Bigg[\big((r+s)P_0 + O_{12}(A(z,r+s),z)\big)\big(1-F_H(A(z,r+s))\big)A_2(z,r+s)\\ &\qquad + \int_0^{A(z,r+s)} P_0\big(1-F_H(x)\big)dx\Bigg] dF_R(r)\\ &= \gamma^f \int_{r_{\min}}^{r_{\max}} \left[\big(rP_0 + O_{12}(A(z,r),z)\big)\big(1-F_H(A(z,r))\big)A_2(z,r) + \int_0^{A(z,r)} P_0\big(1-F_H(x)\big)dx\right] f_R(r-s)dr.\end{aligned}$$
(A.10)

We establish the conditions under which $\frac{\partial^2 C^f}{\partial z \partial s}$ is nonnegative. The only term in (A.8) and (A.10) that can be negative is $O_{12}(A(z,r),z)$. Thus, a sufficient condition for (A.8) and (A.10) to be nonnegative is $rP_0 \ge -O_{12}(A(z,r),z)$. Because $O_1(A(z,r),z) = rP_z = rP_0(1-z)$, this condition is equivalent to $O_{12}(A(z,r),z) \geq -\frac{O_1(A(z,r),z)}{1-z}$. As such, the sufficiency condition given in the Proposition is valid. Even if this sufficiency condition is not satisfied, $\frac{\partial^2 C^f}{\partial z \partial s}$ is typically nonnegative (see example in Figure 2(c)).

When $\frac{\partial^2 C^f}{\partial z \partial s} \geq 0$, the total cost function, $C^f(z,s)$, is supermodular in (z,s). Because we are minimizing a supermodular function, we cannot employ the standard result on maximizing supermodular functions, but need to prove the monotonicity of $z_{\rm DR}^{f*}$ with respect to s directly. Let $s' < s''$, $z' \in \arg\min_z \{C^f(z,s')\}$, and $z'' \in \arg\min_z \{C^f(z,s'')\}$. Suppose $z' < z''$. Then

$$0 \leq C^f(z'',s') - C^f(z',s') \leq C^f(z'',s'') - C^f(z',s'') \leq 0,$$

where the second inequality follows from the supermodularity of $C^f(z,s)$. The above inequalities must hold with equalities. Hence, $z'' \in \arg\min_z \{C^f(z,s')\}$, and $z' \in \arg\min_z \{C^f(z,s'')\}$. This implies that the set $\arg\min_z \{C^f(z,s)\}$ is descending in s. Hence, if the set is a singleton $\{z_{\rm DR}^{f*}\}$, $z_{\rm DR}^{f*}$ decreases in s.

Because $z^{f*} = z_{\rm DR}^{f*}(s)\big|_{s=0}$ in the case of scaling the distribution of R, we have $z_{\rm DR}^{f*} \leq z^{f*}$. ∎

Proof of Lemma 2: We solve $\min_{\alpha \in [0,1]} \left\{O(\alpha h, z) - m\alpha P_z h\right\}$ with $h > 0$. This problem is structurally identical to the problem in Lemma 1. The optimal DR α^{s*} in (14) follows immediately.

Substituting the optimal DR into (13), we can verify the minimum cost expressed in (15). ∎

Proof of Proposition 4: The proof is parallel to the proof for Proposition 2. ∎

Proof of Proposition 5: Because $I(z) \to \infty$ as $z \to 1$, $z_{\rm DR}^{s*} < 1$. Thus, if $z_{\rm DR}^{s*} > 0$, it must be an interior solution and satisfies the first-order condition. It suffices to prove that the first-order derivative of the objective in (16) with respect to z is

$$I'(z) - \gamma^s \bar{c}^s P_0 T + \gamma^s \int_0^\infty \left[\int_0^{B(z,m)} \big(mP_0 + O_{12}(x,z)\big)\big(1 - F_H(x)\big)dx \right] dF_M(m). \qquad \text{(A.11)}$$

Using (15), the problem in (16) can be written as

$$\min_{z \in [0,1)} C^s(z) \equiv I(z) + \gamma^s \left[\bar{G}_b P_z (T - \bar{H}) + \bar{G}_p P_z \bar{H}\right] + \gamma^s \int_0^\infty L^s(z,m) dF_M(m), \qquad \text{(A.12)}$$

where

$$L^s(z,m) = \int_0^{B(z,m)} \big(O(x,z) - mP_z x\big)dF_H(x) + \big(O(B(z,m),z) - mP_z B(z,m)\big)\big(1 - F_H(B(z,m))\big). \qquad \text{(A.13)}$$

Using the techniques of the proof for Proposition 3, we can show that $C^{s\prime}(z)$ is exactly (A.11).

Based on the first order conditions for $z_{\rm DR}^{s*}$ (in (6)) and $z_{\rm DR}^{s*}$ (in (17)), we can see that $z_{\rm DR}^{s*} \leq z^{s*}$

if the following condition holds:

$$\gamma^s \int_0^\infty \left[\int_0^{B(z,m)} \big(mP_0 + O_{12}(x,z)\big)\big(1 - F_H(x)\big)dx \right] dF_M(m) \geq 0. \tag{A.14}$$

Similar to the approach shown in the proof for Proposition 3, a sufficient condition for (A.14) is $-\frac{O_{12}(x,z)}{O_1(x,z)} \leq \frac{1}{1-z}$. ∎

Proof of Proposition 6: Proposition 3 shows that $z^{f*}_{\mathrm{DR}}(r)$ decreases in the DR incentive rate. Thus, the EE gap, $z^{s*}_{\mathrm{DR}} - z^{f*}_{\mathrm{DR}}(r)$, increases in r.

When $r = 0$, we have $A(z,r) = 0$ by definition and $\alpha^{f*} = 0$ by (8), i.e., the firm does not participate in DR. Consequently, the firm decides the EE improvement level as if DR were absent, leading to $z^{f*}_{\mathrm{DR}}(0) = z^{f*}$.

In (11), if $R = r \to 0^+$, then $A(z,r) = O_1^{-1}(rP_z, z) = \inf\{x : O_1(x,z) \geq rP_z\} \to 0$, where the infimum approaches zero because $O_1(x,z)$ is nonnegative and increasing in x. Thus, as $r \to 0^+$, the left-side of (11) approaches $I'(z^{f*}_{\mathrm{DR}}(r))$. Recall from (2) that z^{f*} is determined by $I'(z^{f*}) = \gamma^f cP_0T$. Therefore, $\lim_{r\to 0^+} z^{f*}_{\mathrm{DR}}(r) = z^{f*} = z^{f*}_{\mathrm{DR}}(0)$, implying that $z^{f*}_{\mathrm{DR}}(r)$ is continuous at $r = 0$.

∎

Proof of Proposition 7: The firm's problem in (7) and (10) is repeated here:

$$C(z,h,r) = cP_zT + \min_{\alpha\in[0,1]} \Big\{ O(\alpha h, z) - r\alpha P_z h \Big\},$$

$$\min_{z\in[0,1)} \; I(z) + \gamma^f\, \mathbb{E}\big[C(z,H,R)\big],$$

and the societal problem in (13) and (16) is repeated here:

$$C^s(z,h,g_p,g_b) = g_bP_z(T-h) + g_pP_zh + \min_{\alpha\in[0,1]} \Big\{ O(\alpha h, z) - (g_p - g_b)\alpha P_z h \Big\},$$

$$\min_{z\in[0,1)} \; I(z) + \gamma^s\, \mathbb{E}\big[C^s(z,H,G_p,G_b)\big].$$

When $g_p - g_b > 0$, if we set DR incentive $r = g_p - g_b$, then the inner minimization problem will be the same for the firm and for the society.

When $g_p - g_b \leq 0$, note that the optimal DR level is $\alpha^{s*} = 0$. Zero demand response can be induced by not providing any incentive, i.e., $r = 0$.

Combining the two cases, we see that setting $r = (g_p - g_b)^+$ aligns DR for any given z. Thus, before G_p and G_b realize, we can write the DR incentive as $R = (G_p - G_b)^+$.

If EE incentives are offered to induce the firm to choose the societal-optimal EE improvement level, z^{s*}_{DR}, then the firm would also be participating in the DR as desired by the society. ∎

C. Demand Response Programs with Mandatory Curtailment

In a DR program with mandatory curtailment, the firm pre-commits to a DR participation level for all future DR events. We assume the decision is made simultaneously with the EE investment decision. Such programs include Pacific Gas & Electric's Base Interruptible Program and participation by DR in the Capacity Market of the PJM Interconnection. The firm receives an upfront payment, $\Gamma(\alpha)$, for committing to DR participation level α.

The firm solves the following simultaneous optimization problem:

$$\min_{\alpha\in[0,1],z\in[0,1)} I(z) - \Gamma(\alpha) + \mathbb{E}\left[\sum_{n=1}^{N} \delta_f^n (cP_zT + O(\alpha H, z) - R\alpha P_z H)\right]. \tag{A.15}$$

We assume R and H are independent. While this is a simultaneous optimization over z and α, we may still solve for the optimal α for any z and then optimize over z. Thus, given z, we will seek the α value that minimizes costs. We take out terms in (A.15) that do not depend on α:

$$\min_{\alpha\in[0,1]} -\Gamma(\alpha) + \gamma^f \int_0^{H_{\max}} (O(\alpha x, z) - \bar{R}\alpha P_z x) dF_H(x), \tag{A.16}$$

where $\bar{R} = \mathbb{E}[R]$ denotes the average DR incentive rate.

Taking the derivative with respect to α, we characterize the first-order condition:

$$\gamma^f \int_0^{H_{\max}} xO_1(\alpha x, z) dF_H(x) = \Gamma'(\alpha) + \gamma^f \bar{R} P_z \bar{H}. \tag{A.17}$$

The DR participation level which solves (A.17) is denoted as $\tilde{\alpha}(z, \bar{R})$, as it depends on the EE improvement level, z, and the average DR incentive, $\bar{R}$.

Substituting $\tilde{\alpha}(z, \bar{R})$ into (A.15), we have an optimization only on z:

$$\min_{z\in[0,1)} I(z) - \Gamma(\tilde{\alpha}(z, \bar{R})) + \gamma^f cP_zT - \gamma^f \bar{R}\tilde{\alpha}(z, \bar{R})P_z\bar{H} + \gamma^f \int_0^{H_{\max}} O(\tilde{\alpha}(z, \bar{R})x, z) dF_H(x). \tag{A.18}$$

Taking the derivative with respect to z and simplifying the terms yield the first-order condition for the optimal EE improvement, z:

$$I'(z) - \gamma^f cP_0T + \gamma^f \bar{R}\bar{H}P_0\tilde{\alpha}(z, \bar{R}) + \gamma^f \int_0^{H_{\max}} O_2(\tilde{\alpha}(z, \bar{R})x, z) df_H(x) = 0. \tag{A.19}$$

As in Proposition 3, we show that the firm's choice of EE improvement level decreases when the DR incentive is increased. Using a similar approach as the proof of Proposition 3, we need to show that the objective in (A.18) is supermodular in $(z, \bar{R})$. Taking the derivative of (A.19) with respect to $\bar{R}$, we obtain

$$\gamma^f \bar{H}P_0(\bar{R}\tilde{\alpha}_2(z, \bar{R}) + \tilde{\alpha}(z, \bar{R})) + \gamma^f \int_0^{H_{\max}} O_{12}(\tilde{\alpha}(z, \bar{R})x, z)\tilde{\alpha}_2(z, \bar{R}) x dF_H(x). \tag{A.20}$$

If (A.20) is positive, the objective in (A.18) is supermodular in $(z, \bar{R})$ and we can use the final steps of the proof of Proposition 3 to show that the firm's choice of EE improvement level, z, decreases in the average DR incentive, $\bar{R}$. We may rearrange the terms in (A.20) to get an expression that more closely mirrors (A.8) and (A.10) in the proof of Proposition 3:

$$\bar{H}P_0\tilde{\alpha}(z, \bar{R}) + \int_0^{H_{\max}} x\tilde{\alpha}_2(z, \bar{R})(rP_0 + O_{12}(\tilde{\alpha}(z, r)x, z))dF_H(x) > 0. \tag{A.21}$$

A sufficient condition for (A.21) to hold is $rP_0 > -O_{12}(\tilde{\alpha}(z, r)x, z)$ for $r \in [r_{\min}, r_{\max}]$, i.e., the reduction in marginal overtime costs due to EE is not too large.

While we do not replicate it here, a similar optimization will determine the optimal EE investment from the societal perspective. Again, the EE gap will typically grow with increasing DR incentive.