Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 50-1 (2017) 6761–6767

Learning to Buy (and Sell) Demand Response ⋆

Kia Khezeli * **Weixuan Lin** * **Eilyan Bitar** *

* *School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA. (e-mails:[kk839,wl476,eyb5]@cornell.edu).*

Abstract: We adopt the perspective of an aggregator, which seeks to coordinate its *purchase* of demand reductions from a fixed group of residential electricity customers, with its *sale* of the aggregate demand reduction in a two-settlement wholesale energy market. The aggregator procures reductions in demand by offering its customers a uniform price for reductions in consumption relative to their predetermined baselines. Prior to its realization of the aggregate demand reduction, the aggregator must also determine how much energy to sell into the two-settlement energy market. In the day-ahead market, the aggregator commits to a forward contract, which calls for the delivery of energy in the real-time market. The underlying aggregate demand curve, which relates the aggregate demand reduction to the aggregator's offered price, is assumed to be affine and subject to unobservable, random shocks. Assuming that both the parameters of the demand curve and the distribution of the random shocks are initially unknown to the aggregator, we investigate the extent to which the aggregator might dynamically adapt its DR prices and forward contracts to maximize its expected profit over a window of T days. Specifically, we design a data-driven pricing and contract offering policy that resolves the aggregator's need to learn the unknown demand model with its desire to maximize its cumulative expected profit over time. The proposed pricing policy is proven to exhibit a *regret* over T days that is at most $O(\sqrt{T})$.

Keywords: Demand response, dynamic pricing, online learning, electricity markets.

1. INTRODUCTION

The large scale utilization of demand response (DR) resources has the potential to substantially improve the reliability and efficiency of electric power systems. Accordingly, several state and federal mandates have been established to facilitate the integration of demand response resources into wholesale electricity markets. For example, FERC Order 719 mandates that Independent System Operators (ISOs) permit the direct sale of DR services into wholesale electricity markets (FERC, 2008). As individual residential customers often posses insufficient capacity [1] to participate in such markets directly, there emerges the need for an intermediary, or *aggregator*, with the ability to coordinate the demand response of large numbers of residential customers for direct sale into the wholesale electricity market.

In this paper, we adopt the perspective of an aggregator, which seeks to coordinate its *purchase* of an aggregate demand reduction from a fixed group of residential electricity customers, with its *sale* of the aggregate demand reduction into a two-settlement wholesale energy market. [2] Formally, this amounts to a two-sided optimization problem, which requires the aggregator to balance the cost it incurs in procuring the demand reduction from customers against the revenue it derives from its sale into the wholesale energy market. We develop the problem more formally in what follows.

We consider the setting in which the aggregator purchases demand reductions from customers using a non-discriminatory, price-based mechanism. That is to say, each participating customer is payed for her reduction in electricity demand according to a uniform per-unit energy price determined by the aggregator. Pricing mechanisms of this form fall within the more general category of DR programs that rely on peak time rebates (PTR) as incentives for demand reduction. Prior to its realization of the aggregate demand reduction, the aggregator must also determine how much energy to sell into the two-settlement energy market. In the day-ahead (DA) market, the aggregator commits to a forward contract, which calls for delivery of energy in the real-time (RT) market. If the realized reduction in demand exceeds (falls short of) the forward contract, then the difference is sold (bought) in the RT market. In order to maximize its profit, the aggregator must, therefore, co-optimize the DR price it offers its customers with the forward contract that it commits to in the wholesale energy market.

⋆ This work was supported in part by NSF grant ECCS-1351621, NSF grant CNS-1239178, NSF grant IIP- 1632124, US DoE under the CERTS initiative, and the Simons Institute for the Theory of Computing.

[1] For example, the Proxy Demand Resource (PDR) program currently being operated by the California ISO has a minimum curtailment capacity requirement of 100 kW (Wolak et al., 2009).

[2] We note that a measurable reduction in demand is equivalent to an increase in supply.

10.1016/j.ifacol.2017.08.1193

There are a myriad of challenges that the aggregator faces in the deployment of such programs. The most basic challenge is the prediction of how customers will adjust their aggregate demand in response to different DR prices, i.e., the aggregate demand curve. If the offered price is too low, consumers may be unwilling to curtail their demand; if the offered price is too high, the aggregator pays too much and gets more reduction than is needed. As the aggregator is initially ignorant to customers' aggregate demand curve, the aggregator must attempt to learn a model of customer behavior over time through repeated observations of demand reductions in response to its offered DR prices. Simultaneously, the aggregator must jointly adjust its DR prices and forward contract offerings in such a manner as to facilitate profit maximization over time. As we will later show, such tasks are intimately related, and give rise to a trade-off between the need to *learn* (explore) and *earn* (exploit).

Contribution and Related Work: We study the setting in which the aggregator is faced with an aggregate demand curve that is affine in price, and subject to unobservable, additive random shocks. We assume that both the parameters of the demand curve and the probability distribution of the random shocks are fixed, and *initially unknown* to the aggregator. Faced with such ignorance, we explore the extent to which the aggregator might dynamically adapt its posted DR prices and offered contracts to maximize its expected profit over a time frame of T days. Specifically, we design a causal pricing and contract offering policy that resolves the aggregator's need to learn the unknown demand model with its desire to maximize its cumulative expected profit over time. The proposed pricing policy is proven to exhibit *regret* (relative to an oracle) over T days that is at most $O(\sqrt{T})$. In addition, the proposed policy generates a sequence of posted DR prices and forward contracts that converge to the oracle optimal DR price and forward contract in the mean square sense.

The literature – as it relates to the problem of co-optimizing an aggregator's decisions in both the retail and wholesale electricity markets – is sparse. Campaigne and Oren (2015) consider a market model that is perhaps closest in nature to the one considered in this paper. They adopt a mechanism design approach to eliciting demand response, where customers are rationed and remunerated according to their reported types. A related line of literature includes (Chao, 2012) and (Crampes and Léautier, 2015). In this paper, we take a posted price approach to the procurement of demand response. This is in sharp contrast to the mechanism design approach, as it gives rise to the need to learn customers' types (i.e., demand function) over time.

Organization: The remainder of the paper is organized as follows. In Section 2, we formulate the aggregator's profit maximization problem. In Section 3, we propose a recursive estimation scheme to learn the unknown demand model. In Section 4, we propose a joint pricing and contract offering policy for the aggregator, and provide a theoretical analysis of the regret incurred by the policy. In Section 5, we illustrate the performance of our proposed policy with a numerical example. All mathematical proofs are omitted in this version of the paper due to space constraints.

2. MODEL

We adopt the perspective of an aggregator who seeks to purchase demand reductions from a fixed group of N customers for sale into a two-settlement energy market. The market is assumed to repeat over multiple time periods (e.g., days) indexed by $t = 1, 2, \ldots$. The actions taken by the aggregator and their timing are specified in the following subsections.

2.1 Two-Settlement Market Model

At the beginning of each time period t, the aggregator commits to a forward contract for energy in the day-ahead (DA) market in the amount of Q_t (kWh). The forward contract is remunerated at the *DA energy price*. The forward contract calls for delivery in the real-time (RT) market. If the energy delivered by the aggregator (i.e., demand reduction) falls short of the forward contract, the aggregator must purchase the shortfall in the RT market at the *shortage price*. If the energy delivered exceeds the forward contract, the aggregator must sell the excess supply in the RT market at the *overage price*. We denote the wholesale energy prices ($/kWh) by

- π, DA energy price,
- π_-, RT shortage price,
- π_+, RT overage price.

Although we assume throughout the paper that the wholesale energy prices are fixed and known across time, all results stated in this paper can be generalized to accommodate the more general setting in which the wholesale energy prices exhibit known variation with time. We also assume that the wholesale energy prices satisfy $\pi > 0$ and $\pi_+ < \pi < \pi_-$. Such assumption serves to facilitate clarity of exposition and analysis in the sequel, as it preserves concavity of the aggregator's expected profit function (2).

2.2 Demand Response Model

In order to meet its forward contract commitment Q_t, the aggregator must elicit an aggregate reduction in demand from its customers. It does so by broadcasting a uniform DR price $p_t \geq 0$, to which each customer i responds with a reduction in demand in the amount of D_{it} (kWh), thereby entitling each customer i to receive a payment of $p_t D_{it}$. Implicit in this model is the assumption that each customer's reduction in demand is measured against a predetermined baseline.

We model the response of each customer i to the posted price p_t at time t according to the *affine* function

$$D_{it} = a_i p_t + b_i + \varepsilon_{it}, \quad \text{for } i = 1, \ldots, N,$$

where $a_i \in \mathbb{R}$ and $b_i \in \mathbb{R}$ are the demand model parameters, and ε_{it} is an unobservable demand shock, which we model as a zero-mean random variable. We assume that both the model parameters a_i and b_i, and the distribution function of the demand shock are initially

unknown to the aggregator. Clearly, the aggregate demand reduction $D_t := \sum_{i=1}^N D_{it}$ satisfies the affine relationship

$$D_t = ap_t + b + \varepsilon_t, \tag{1}$$

where the aggregate model parameters and shock are defined as $a := \sum_{i=1}^N a_i$, $b := \sum_{i=1}^N b_i$, and $\varepsilon_t := \sum_{i=1}^N \varepsilon_{it}$, respectively. We denote by $\theta := (a, b)$ the demand model parameter.

We assume throughout the paper that $a \in [\underline{a}, \overline{a}]$ and $b \in [0, \overline{b}]$ where $\underline{a}$, $\overline{a}$, and $\overline{b}$ are known and satisfy $0 < \underline{a} \leq \overline{a} < \infty$ and $0 \leq \overline{b} < \infty$. Such assumptions are natural, as they ensure a bounded and positive price elasticity of aggregate demand, and that reductions in aggregate demand are guaranteed to be nonnegative in the absence of demand shocks. We also assume that the sequence of aggregate demand shocks $\{\varepsilon_t\}$ are independent and identically distributed (IID) random variables, in addition to the following technical assumption.

Assumption 1. The aggregate demand shock ε_t takes values in the interval $[\underline{\varepsilon}, \overline{\varepsilon}]$. Moreover, its cumulative distribution function F is bi-Lipschitz over this range. Namely, there exists a real constant $L \geq 1$, such that for all $x, y \in [\underline{\varepsilon}, \overline{\varepsilon}]$, it holds that

$$\frac{1}{L}|x - y| \leq |F(x) - F(y)| \leq L|x - y|.$$

The assumption that the aggregate demand shock takes bounded values is natural, given the physical limitation on the range of values that demand can take. We also note that the aggregator does not require explicit knowledge of the parameters specified in Assumption 1.

2.3 Aggregator Profit

The expected profit derived by the aggregator during period t under a fixed price p_t and forward contract Q_t is given by

$$\begin{aligned} &r(p_t, Q_t) \\ &\quad := \pi Q_t + \mathbb{E}\left[\pi_+[D_t - Q_t]^+ - \pi_-[Q_t - D_t]^+ - p_t D_t\right], \end{aligned} \tag{2}$$

where the expectation is taken with respect to the distribution on the random shock ε_t, and $[x]^+ := \max\{0, x\}$ for all $x \in \mathbb{R}$. It is not difficult to show that the expected profit criterion (2) is concave in its arguments (p_t, Q_t) given the assumptions stated in this paper thus far.

We define the *oracle optimal price and contract* as

$$(p^*, Q^*) := \operatorname{argmax}\{r(p, Q) : (p, Q) \in \mathbb{R}^2\}.$$

That is to say, (p^*, Q^*) denote the DR price and forward contract, which jointly maximize the aggregator's expected profit given perfect knowledge of the demand model. Note that oracle optimal price and contract are time-invariant, as the wholesale energy prices and demand model are time invariant. Their closed-form expressions are given in the following lemma.

Lemma 1. (Oracle Optimal Policy). The oracle optimal price p^* and contract Q^* are given by

$$p^* = \frac{1}{2}\left(\pi - \frac{b}{a}\right), \tag{3}$$

$$Q^* = ap^* + b + F^{-1}(\alpha), \tag{4}$$

where $\alpha := (\pi - \pi_+)/(\pi_- - \pi_+)$.

Here, $F^{-1}(\alpha) := \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$ denotes the α-quantile of the random shock ε_t. We are guaranteed that $\alpha \in [0, 1]$, because of the assumption that $\pi_+ < \pi < \pi_-$.

We define the *oracle optimal profit* accumulated over T time periods as

$$R^*(T) := \sum_{t=1}^T r(p^*, Q^*).$$

We employ the term *oracle*, as $R^*(T)$ equals the maximum expected profit that an aggregator might derive over T times periods if it had perfect knowledge of the demand model.

2.4 Policy Design and Regret

We consider the scenario in which the aggregator knows neither the demand model parameter $\theta = (a, b)$ nor the aggregate shock distribution F at the outset. Accordingly, the aggregator must endeavor to learn these features from the data it collects over time, e.g., measurements of aggregate demand reductions in response to its posted DR prices. At the same time, the aggregator must dynamically adapt its sequence of posted DR prices (and forward contract offerings) to improve its profit over time. In what immediately follows, we describe the space of feasible policies that the aggregator might use to guide its adaptation of DR prices $\{p_t\}$ and contracts $\{Q_t\}$ over time.

Prior to its determination of the price p_t and the contract Q_t at time t, the aggregator has access to the entire history of prices, contract offerings, and aggregate demand reductions, up to and including time period $t - 1$. We define a *feasible policy* as an infinite sequence of functions $\gamma := ((p_1, Q_1), (p_2, Q_2), \ldots)$, where each function in the sequence is allowed to depend only on the past data available until that point in time. More formally, we require that the functions (p_t, Q_t) be measurable according to the σ-algebra generated by the history of prices, offered contracts, and demand observations, i.e.,

$$(p_1, \ldots, p_{t-1}, Q_1, \ldots, Q_{t-1}, D_1, \ldots, D_{t-1})$$

for all time periods $t \geq 2$. For $t = 1$, we require that (p_1, Q_1) be a pair of deterministic constants.

The *expected profit* generated by a feasible policy γ over T time periods is defined as

$$R^\gamma(T) := \mathbb{E}^\gamma\left[\sum_{t=1}^T r(p_t, Q_t)\right], \tag{5}$$

where the expectation is taken with respect to the demand model (1) under the policy γ. We measure the performance of a feasible policy γ over T time periods according to the *T-period regret*:

$$\Delta^\gamma(T) := R^*(T) - R^\gamma(T).$$

The T-period regret incurred by a feasible policy equals the difference between the oracle optimal profit and the expected profit incurred by that policy over T time periods. Clearly, policies that produce low regret are preferred, as the oracle optimal profit is an upper bound on the expected profit achievable by any feasible policy. Accordingly, we seek the design of policies whose T-period regret grows sublinearly with the horizon T. Such policies are said to have *no-regret*, as their average regret $(1/T) \cdot \Delta^\gamma(T)$ is

guaranteed to vanish asymptotically. More formally, we have the following definition.

Definition 1. (No-Regret Policy). A feasible policy γ is said to have *no-regret* if $\lim_{T\to\infty} \Delta^\gamma(T)/T = 0$.

The following result establishes an upper bound on the T-period regret in terms of squared pricing and contract errors relative to their oracle optimal counterparts. Lemma 2 will prove useful to the derivation of our main results.

Lemma 2. The T-period regret incurred by any feasible policy γ is upper bounded by

$$\Delta^\gamma(T) \le a \sum_{t=1}^{T} \mathbb{E}^\gamma \left[(p_t - p^*)^2\right] + L(\pi_- - \pi_+) \sum_{t=1}^{T} \mathbb{E}^\gamma \left[\left(Q_t - Q^* - a(p_t - p^*)\right)^2\right], \quad (6)$$

where (p^*, Q^*) denote the oracle optimal price and contract.

Lemma 2 reveals that convergence of the posted prices $\{p_t\}$ and offered contracts $\{Q_t\}$ to the oracle optimal price p^* and contract Q^* in the mean square sense, respectively, will prove essential to the design of policies that exhibit no-regret. In the following section, we introduce a simple (least-squares) method to learning the demand model that will facilitate the design of such policies.

3. DEMAND MODEL LEARNING

In this section, we propose a simple approach to learning the demand model from data using the method of least squares estimation.

3.1 Parameter Estimation

We define the *least squares estimator* (LSE) of the parameter θ, given the history of past prices and demand observations at time period t as

$$\theta_t := \arg\min \left\{ \sum_{k=1}^{t-1} \left(D_k - (\vartheta_1 p_k + \vartheta_2)\right)^2 : (\vartheta_1, \vartheta_2) \in \mathbb{R}^2 \right\},$$

for time periods $t = 2, 3, \ldots$. It is straightforward to show that

$$\theta_t = \mathscr{J}_{t-1}^{-1} \left(\sum_{k=1}^{t-1} \begin{bmatrix} p_k \\ 1 \end{bmatrix} D_k \right), \quad (7)$$

assuming that the indicated inverse exists. The matrix $\mathscr{J}_t$ is defined as

$$\mathscr{J}_t := \sum_{k=1}^{t} \begin{bmatrix} p_k \\ 1 \end{bmatrix} \begin{bmatrix} p_k \\ 1 \end{bmatrix}^\top.$$

Its inverse is given by

$$\mathscr{J}_t^{-1} = J_t^{-1} \left(\frac{1}{t} \sum_{k=1}^{t} \begin{bmatrix} -1 \\ p_k \end{bmatrix} \begin{bmatrix} -1 \\ p_k \end{bmatrix}^\top \right), \quad (8)$$

where $J_t := \sum_{k=1}^{t} (p_k - \bar{p}_t)^2$, and $\bar{p}_t := (1/t) \sum_{k=1}^{t} p_k$. The parameter estimation error that results under the LSE (7) can be expressed as

$$\theta_t - \theta = \mathscr{J}_{t-1}^{-1} \left(\sum_{k=1}^{t-1} \begin{bmatrix} p_k \\ 1 \end{bmatrix} \varepsilon_k \right). \quad (9)$$

Remark 1. (The Role of Price Exploration) The expression for the parameter estimation error in (9) implies a sufficient condition on the sequence of prices, which guarantees consistency of the LSE. Namely, the parameter estimation error converges to zero in probability if the sequence of prices are such that J_t grows unbounded with time, almost surely. Consequently, a policy that guarantees sufficient *price exploration*, i.e., persistent variation in the sequence of prices, results in consistency of the parameter estimates. In Section 4, we propose a policy that generates enough variation in the sequence of prices such that J_t is guaranteed to be at least $O(\sqrt{t})$.

Recalling our previous assumption that the unknown parameter θ belongs to a closed and compact set given by $\Theta := [\underline{a}, \overline{a}] \times [0, \overline{b}]$, one can improve upon the LSE (7) by projecting θ_t onto the set Θ. More precisely, define the *truncated least squares estimator* as

$$\widehat{\theta}_t := \arg\min \left\{ \|\vartheta - \theta_t\|_2 : \vartheta \in \Theta \right\}. \quad (10)$$

It clearly holds that $\|\widehat{\theta}_t - \theta\| \le \|\theta_t - \theta\|$.

3.2 Quantile Estimation

We propose an approach to the recursive estimation of the unknown quantile function using the residuals generated by the truncated LSE (10). At each time period t, define the sequence of *residuals* associated with the estimator $\widehat{\theta}_t$ as

$$\widehat{\varepsilon}_{k,t} := D_k - (\widehat{a}_t p_k + \widehat{b}_t), \quad \text{for } k = 1, \ldots, t. \quad (11)$$

Define their *empirical distribution* as

$$\widehat{F}_t(x) := \frac{1}{t} \sum_{k=1}^{t} \mathbb{1}\{\widehat{\varepsilon}_{k,t} \le x\},$$

and their corresponding *empirical quantile function* as $\widehat{F}_t^{-1}(\alpha) := \inf\{x \in \mathbb{R} : \widehat{F}_t(x) \ge \alpha\}$. It will prove useful to the subsequent analyses to express the empirical quantile function in terms of the order statistics associated with the sequence of residuals. The *order statistics* associated with the sequence $\widehat{\varepsilon}_{1,t}, \ldots, \widehat{\varepsilon}_{t,t}$ are defined as a permutation of the sequence denoted by $\widehat{\varepsilon}_{(1),t}, \ldots, \widehat{\varepsilon}_{(t),t}$, where

$$\widehat{\varepsilon}_{(1),t} \le \widehat{\varepsilon}_{(2),t} \le \ldots \le \widehat{\varepsilon}_{(t),t}.$$

With the order statistics of the residuals in hand, one can express the empirical quantile function as

$$\widehat{F}_t^{-1}(\alpha) = \widehat{\varepsilon}_{(i),t}, \quad (12)$$

where i is the unique index such that $i - 1 < t\alpha \le i$, i.e., $i = \lceil t\alpha \rceil$. Using Equation (12), the quantile estimation error can be linked to the parameter estimation error via the following inequality,

$$|\widehat{F}_t^{-1}(\alpha) - F^{-1}(\alpha)| \le |F_t^{-1}(\alpha) - F^{-1}(\alpha)| + \left(1 + |p_{(i)}|\right) \|\widehat{\theta}_t - \theta\|_1, \quad (13)$$

where $F_t^{-1}(\alpha)$ is defined as the empirical quantile function associated with the sequence of demand shocks $\varepsilon_1, \ldots, \varepsilon_t$.

It follows from the inequality in (13) that consistency of the quantile estimator (12) depends on consistency of both the parameter estimator $\widehat{\theta}_t$ and the empirical quantile function $F_t^{-1}(\alpha)$. We establish consistency of the parameter estimator under our proposed policy in Lemma 3. Clearly, consistency of the empirical quantile function

$F_t^{-1}(\alpha)$ does not depend on the particular policy being used. In Proposition 1, we establish a bound on the rate at which the sequence $\{F_t^{-1}(\alpha)\}$ converges to $F^{-1}(\alpha)$ in probability.

Proposition 1. There exists a finite positive constant μ_1 such that

$$\mathbb{P}\{|F_t^{-1}(\alpha) - F^{-1}(\alpha)| > \epsilon\} \leq 2\exp(-\mu_1\epsilon^2 t) \tag{14}$$

for all $\epsilon > 0$ and $t \geq 2$.

4. BUYING AND SELLING WITH NO-REGRET

In this section, we build on the approach to demand model learning outlined in Section 3 to construct a pricing and contract offering policy, which is guaranteed to exhibit *no-regret*.

4.1 Myopic Policy

We first introduce a natural approach to pricing and contract offering, which combines the model estimation scheme outlined in Section 3 with a *myopic* approach to pricing and contract offering. That is to say, at each time period t, the aggregator estimates the demand model parameters and quantile function according to (10) and (12), respectively, and sets the price and forward contract according to

$$\widehat{p}_t = \frac{1}{2}\left(\pi - \frac{\widehat{b}_t}{\widehat{a}_t}\right), \tag{15}$$

$$\widehat{Q}_t = \widehat{a}_t\widehat{p}_t + \widehat{b}_t + \widehat{F}_t^{-1}(\alpha). \tag{16}$$

Under this myopic policy, the aggregator treats its demand model estimates in each period as if they were correct, and ignores the impact of its choice of price on its ability to accurately estimate the demand model in future time periods. As discussed in Remark 1, consistency of the parameter estimator is reliant upon sufficient variation in the underlying sequence of prices. However, under the myopic policy the sequence of prices may converge prematurely to a fixed price (that is different from the oracle optimal price). As a consequence, the sequence of parameter estimates may also converge to values different from the true model parameter. This phenomenon, also known as *incomplete learning*, is well-documented in the revenue management literature – see, for example, (Lai and Robbins, 1982; Keskin and Zeevi, 2014). In Section 5, we conduct a numerical case study that appears to indicate the occurrence of incomplete learning under the myopic policy – see, for example, Figures 1(h) and 1(i).

4.2 Perturbed Myopic Policy

To guarantee sufficient price exploration, we propose a policy that is similar in structure to a policy first introduced in (Khezeli and Bitar, 2016a,b). We refer to this policy as the *perturbed myopic policy*. The policy forces price exploration by adding a perturbation (of appropriate magnitude) to the myopic price at every other time step. More precisely, we define the perturbed myopic policy as

$$p_t = \begin{cases} \widehat{p}_t, & t \text{ odd}, \\ \widehat{p}_{t-1} + \rho t^{-1/4}, & t \text{ even}, \end{cases} \tag{17}$$

$$Q_t = \widehat{a}_t p_t + \widehat{b}_t + \widehat{F}_t^{-1}(\alpha), \tag{18}$$

where $\rho \geq 0$ is a user specified constant that we allow to be arbitrary in this paper.[3] There exists a natural trade-off in setting the price perturbation. On the one hand, the perturbations should decay at a rate that is slow enough to generate sufficient price exploration required to ensure consistent parameter estimation. On the other hand, the perturbations should decay at a rate that is fast enough to guarantee a sublinear growth rate of regret. In fact, it can be shown that among all polynomial functions of t, the optimal choice of the price perturbation (up to a multiplicative constant), which minimizes the asymptotic order of our upper bound on regret is given by $t^{-1/4}$.

Recall the upper bound on the T-period regret established in Lemma 2. Upon examination of the inequality in (6), it becomes apparent that it suffices to bound the rate at which the squared pricing and contract errors accumulate under the perturbed myopic policy, in order to upper bound the rate at which regret accumulates. We do so by first relating the pricing and contract errors to the parameter estimation error. We then derive a bound on the rate at which the parameter estimation error converges to zero in probability under the perturbed myopic policy.

By combining Equations (15) and (3), we can upper bound the pricing error by

$$|\widehat{p}_t - p^*| \leq k_1\|\widehat{\theta}_t - \theta\|_1, \tag{19}$$

where $k_1 := (a+b)/(2\underline{a}a)$. Similarly, by combining Equations (16) and (4) with the inequality in (13), we can upper bound the contract error by

$$|\widehat{Q}_t - Q^*| \leq k_2\|\widehat{\theta}_t - \theta\|_1 + |F_t^{-1}(\alpha) - F^{-1}(\alpha)|, \tag{20}$$

where $k_2 := (2\overline{p} + \pi + 3)/2$ and $\overline{p} := \rho + (\pi - \underline{b}/\overline{a})/2$.

Proposition 1 establishes a bound on the rate at which the last term in (20) converges to zero in probability. The following Lemma establishes a bound on the rate at which the sequence of parameter estimates $\{\widehat{\theta}_t\}$ (generated under the perturbed myopic policy) converges to the true parameter θ in probability.

Lemma 3. (Consistent Parameter Estimation). There exist finite positive constants μ_2 and μ_3 such that, under the perturbed myopic policy (17) and (18),

$$\mathbb{P}\{\|\widehat{\theta}_t - \theta\|_1 > \epsilon\} \leq 2\exp(-\mu_2\epsilon^2\sqrt{t}) + 2\exp(-\mu_3\epsilon^2 t)$$

for all $\epsilon > 0$ and $t \geq 2$.

The following Theorem establishes an $O(\sqrt{T})$ upper bound on the T-period regret incurred by the perturbed myopic policy.

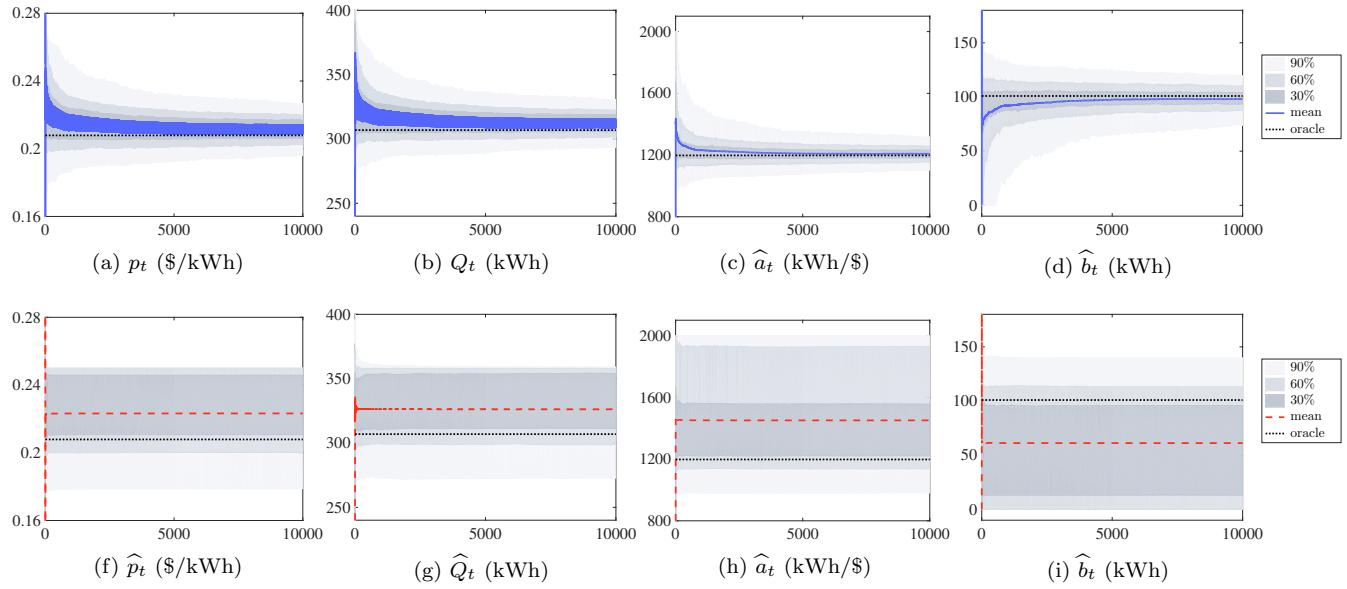
Theorem 1. (Sub-linear Regret). There exists a finite positive constant K such that, under the perturbed myopic policy (17) and (18), the T-period regret is bounded by

$$\Delta(T) \leq K\sqrt{T}$$

for all $T \geq 2$.

As part of the proof of Theorem 1, we establish that the sequences of posted prices $\{p_t\}$ and contracts $\{Q_t\}$ generated by the perturbed myopic policy converge to the oracle optimal price p^* and contract Q^* in the mean

[3] Note that ρ plays a role in determining the finite-time behavior of the perturbed myopic policy. Nevertheless, the asymptotic order of regret incurred by the policy remains the same for any choice of $\rho > 0$.



(a) p_t (\$/kWh) (b) Q_t (kWh) (c) $\widehat{a}_t$ (kWh/\$) (d) $\widehat{b}_t$ (kWh)

(f) $\widehat{p}_t$ (\$/kWh) (g) $\widehat{Q}_t$ (kWh) (h) $\widehat{a}_t$ (kWh/\$) (i) $\widehat{b}_t$ (kWh)

Fig. 1. Sequences of prices, contract offerings, and paramter estimates generated by the perturbed myopic policy (top) and the myopic policy (bottom), compared against their oracle policy counterparts. The shaded area represents their empirical confidence interval estimated using 500 independent realizations of the sequence of demand shocks.

square sense, respectively. We also remark that Chen et al. (2014) consider a similar setting, which entails the online control of a dynamic inventory system through pricing and ordering decisions. They consider a different class of policy designs, and similarly establish an $O(\sqrt{T})$ upper bound on the order of regret for the class of policies they consider.

5. CASE STUDY

In this section, we compare the performance of the myopic policy against the perturbed myopic policy (with $\rho = 0.05$) over a time horizon of $T = 10^4$ periods. We assume that there are $N = 10^4$ customers participating in the DR program. For each customer i, we select a_i uniformly at random from the interval $[0.04, 0.20]$, and independently select b_i according to an exponential distribution (with mean equal to 0.01) truncated over the interval $[0, 0.1]$. [4] Parameters are drawn independently across customers. For each customer i, we let the demand shock have a normal distribution with zero-mean and standard deviation equal to 0.5, truncated over the interval $[-2, 2]$. We set the DA energy price, the RT shortage price, and the RT overage price to $\pi = 0.5$, $\pi_- = 1.7$, and $\pi_+ = 0.2$ (\$/kWh), respectively. Finally, we estimate the mean values and confidence intervals associated with price, contract, and parameter estimate trajectories using 500 independent realizations of the experiment.

5.1 Discussion

Figure 1(f) illustrates an apparent lack of exploration in the sequence of posted prices generated by the myopic policy. That is to say, the myopic price sequence rapidly converges to a fixed value, which on average substantially differs from the oracle optimal price. The same is true for the sequence of forward contracts generated by the myopic policy, as can be seen from Figure 1(g). The premature convergence of the myopic price sequence, in turn, leads to incomplete learning, as is depicted in Figures 1(h) and 1(i). As a consequence, the T-period regret incurred by the myopic policy grows linearly in T, as shown in Figure 2.

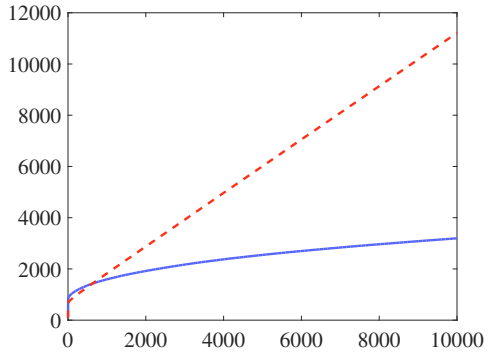


Fig. 2. A plot of the T-period regret incurred by the perturbed myopic policy (——) compared to the T-period regret incurred by the myopic policy (- - -).

On the other hand, the persistent variation in the sequence of prices generated by the perturbed myopic policy induces parameter estimates, which asymptotically converge to the true parameter values, as can be seen from Figures 1(c) and 1(d). In Figures 1(a) and 1(b), one can observe that the confidence intervals associated with the posted price and contract sequences generated by the perturbed myopic policy shrink to the optimal oracle values over time. This provides empirical evidence supporting our theoretical claim that the sequences of prices and contracts generated by the perturbed myopic policy converge to their oracle optimal values in probability.

[4] This range of parameter values is consistent with the range of demand price elasticities observed in several real-time pricing programs operated in the United States, (DoE, 2006; Faruqui and Sergici, 2010).

6. CONCLUSION

In this paper, we study the problem of optimizing the expected profit of an aggregator. The aggregator purchases energy in the form of demand reductions from a fixed group of residential customers, and sells the (uncertain) aggregate demand reduction in a two-settlement wholesale electricity market. The customers' aggregate demand function is assumed to be affine in price (with unknown parameters) and subject to unobservable, additive random shocks (with unknown distribution). We propose a data-driven policy for setting DR prices and forward contract offerings. We show that the proposed policy is consistent, meaning that the sequences of prices and contracts that it generates converge to the oracle optimal price and contract in the mean square sense, respectively. Moreover, we show that the regret incurred by the proposed policy over T time periods is no more than $O(\sqrt{T})$.

Although the perturbed myopic policy that we propose yields a regret with a sublinear growth rate in the time horizon T, its finite-time performance may leave something to be desired. That is to say, the profit loss over finite time horizons may be quite large in practice. Thus, as a direction for future research, it would be of interest to explore the design of policies with improved finite-time performance guarantees.

REFERENCES

Campaigne, C. and Oren, S.S. (2015). Firming renewable power with demand response: an end-to-end aggregator business model. *Journal of Regulatory Economics*, 1–37.

Chao, H.p. (2012). Competitive electricity markets with consumer subscription service in a smart grid. *Journal of Regulatory Economics*, 41(1), 155–180.

Chen, B., Chao, X., and Ahn, H.S. (2014). Coordinating pricing and inventory replenishment with nonparametric demand learning. *Available at SSRN 2694633*.

Crampes, C. and Léautier, T.O. (2015). Demand response in adjustment markets for electricity. *Journal of Regulatory Economics*, 48(2), 169–193.

DoE (2006). Benefits of demand response in electricity markets and recommendations for achieving them. *US Dept. Energy, Washington, DC, USA, Tech. Rep.*

Faruqui, A. and Sergici, S. (2010). Household response to dynamic pricing of electricity: a survey of 15 experiments. *Journal of Regulatory Economics*, 38(2), 193–225.

FERC (2008). 719, Wholesale competition in regions with organized electric markets. *Federal Energy Regulatory Commission*.

Keskin, N.B. and Zeevi, A. (2014). Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5), 1142–1167.

Khezeli, K. and Bitar, E. (2016a). Data-driven pricing of demand response. In *Smart Grid Communications (SmartGridComm), 2016 IEEE International Conference on*, 224–229. IEEE.

Khezeli, K. and Bitar, E. (2016b). Risk-sensitive learning and pricing for demand response. *arXiv preprint arXiv:1611.07098*.

Lai, T. and Robbins, H. (1982). Iterated least squares in multiperiod control. *Advances in Applied Mathematics*, 3(1), 50–73.

Wolak, F.A., Bushnell, J., and Hobbs, B.F. (2009). The California ISO's proxy demand resource (PDR) proposal. *Market Surveillance Committee of the California ISO May*, 1.